

INVESTIGATIONS INTO HOW BACTERIA INFLUENCE NUTRIENT
AVAILABILITY IN THEIR ENVIRONMENT

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

David Robert Sannino

December 2017

© 2017 David Robert Sannino

INVESTIGATIONS INTO HOW BACTERIA INFLUENCE NUTRIENT AVAILABILITY IN THEIR ENVIRONMENT

David Robert Sannino, Ph. D.

Cornell University 2017

Bacteria shape many of their interactions with other organisms through the manipulation of nutrient availability, whether through cooperation by providing nutrients, or through competition for nutrients. Thiamin is an essential vitamin necessary for all life, however, how bacteria shape the ecological interactions between organisms for this nutrient is not well understood. We employed a *Drosophila melanogaster*-microbiota model utilizing a chemically defined diet to understand the interaction between how the host is influenced by the microbiota's interaction with the dietary component thiamin. We found that the *Drosophila melanogaster* microbiota provisions thiamin to its host in a low thiamin environment. This provision rescued development of *Drosophila* on a no thiamin diet, as axenic flies were unable to develop on this diet. Our study was a clear demonstration supporting the long standing hypothesis that animal microbiotas function to provision thiamin and other vitamins to their host. A small subset of bacteria produce the enzyme thiaminase I, which degrades thiamin to its two moieties. The biological function of this enzyme is not understood. We used a genomic approach to investigate a potential function of this enzyme and found that it is located in a conserved operon in three thiaminase I producing *Paenibacillus* species, with other genes involved in thiamin salvage and production of the thiamin antimetabolite bacimethrin, suggesting it may play a role in thiamin salvage and competition for this nutrient. We further

investigated the biological role of thiaminase I using *Burkholderia thailandensis*, where we generated thiamin auxotrophs. We found that the enzyme plays a role in thiamin salvage as it allows for auxotrophic strains to grow in media conditions when strains lacking thiaminase I cannot, as it recycles precursors from thiamin and certain analogs. Using a genomic approach, we also investigated the biosynthetic and metabolic potential of the unique, giant bacterial intestinal symbiont ‘*Candidatus* Epulopiscium viviparous’ of the surgeonfish *Naso tonganus*. We found that this bacterium’s genome is enriched for carbohydrate metabolism, as it has the potential to degrade a vast array of carbohydrates present in its host’s diet, allowing them to supply assimilable nutrients, vitamins, and protein to their host.

BIOGRAPHICAL SKETCH

Dave Sannino was born on October 24, 1987, and grew up in Millstone Township, NJ. He attended Rutgers University where he graduated Summa Cum Laude. While at Rutgers University, he worked as an undergraduate researcher in Dr. Elisabetta Bini's laboratory studying the extremophilic organisms including the archaeon *Sulfolobus solfataricus*. In August 2011, he began his graduate career in the Department of Microbiology at Cornell University, ultimately finding himself in Dr. Esther Angert's laboratory, where he began studying the biology and genomics of *Epulopiscium* spp. as well as the bacterial produced enzyme thiaminase I. Throughout his graduate career he was a teaching assistant for multiple microbiology courses and received the CALS Outstanding TA Award for microbiology in 2015, for his TAsip in general microbiology. In his spare time, Dave enjoys playing hockey, drawing, and music.

ACKNOWLEDGMENTS

I am deeply grateful to my advisor Esther Angert for all of her guidance, assistance, hard work, and mentorship, as well as for giving me the freedom to explore multiple projects, systems, and techniques throughout my graduate career. Also for being very patient with me and not giving up on me particularly after my rough A-exam. I am grateful to my minor advisor Nicolas Buchon for taking me into your lab, giving me advice, expertise, and mentorship. I would not have been able to complete my PhD without you. I thank you to my lab-mate Francine Arroyo for all the assistance, the advice, hard work, genomic expertise, and willingness to always help me no matter what. I would not have been able to complete the *Epulopiscium* chapter without her and I am indebted to her for that. I would like to thank Cliff Kraft for all his insight, discussions, and conversations in regards to the thiaminase I project. I would like to thank Chuck Pepe-Ranney for his genomics wizardry, willingness to always help and answer questions, and great insight. Without him, I would still be lost trying to figure out how to assemble the type B genome. I would like to thank Katie Edwards for all of her expertise on thiamin, for supplying me with multiple analogs, for all her hard work and late nights performing the thiamin assays, and for always taking the time to thoughtfully answer every question I have about thiamin. I thank my other committee member John Helmann for his insight and helpful discussions. I would like to thank Adam Dobson for his mentorship, statistical help, fly insight, friendship, and all the good nights at the Chapter House, Statler, and Westy. I am thankful to the past Angert lab members; Betsy Hutchison and Dave Miller for getting me acquainted with epulos, Jen Fownes for all her help with thiaminase I, and Verena Carvalho-Salman, Bobby Loren, Abe Francis, Jenny Zhang, and Zanah Francis. I would like to thank the Buchon lab, in particular Alessandro Bonfini, Phil Houtz, Yoni Revah, Xi Liu, and Xiaoli Bing for their help with fly work. I thank Anthony Hay and Steve Winans for the 2 great semesters TAing 2900, as well as Steve Zinder, Barb Eaglesham, and Sue Merkel for all their mentorship with teaching. I would like to thank Joe Peters and Ahmed Gaballa for their insight, as well as Pete Newell for his insight and mentorship in my early stages of my grad career. I would like to thank Françoise Vermeulen at the Cornell Statistical Consulting Unit and Qi Sun

from the Cornell Institute of Biotechnology for all their assistance.

I would like to thank my parents, Robert and Irene Sannino for their love and support throughout these 6+ years, as well as my brother Doug and my sister Liane. I would not have been able to last the 6+ years without you. I would like to thank my cohort, Olya Lastovetsky, Chantal Koechli, and Evgeniya Nazarova for the fun and friendship. I am thankful to Jared, Schwenke, Ginny, Marc, Bryan, Sara, Calum, Sören, Daniel, Dane, and everyone else for the friendship and good times. I am thankful to Margarita Kazakova and Rasa Žalakevičiūtė as well for their love and support when we were together. I would like to thank James Orcutt for helping to grow the hockey community in Ithaca, all the hockey people (except Pure Romance), a lot of my best times in Ithaca were on Cass Park and the Rink ice. Last but certainly not least, I have to thank Andrew, Barb, Bondar, Ceante, Leland, Mike, Pete, and Tom, for all the support and everything else.

TABLE OF CONTENTS

BIOGRAPHICAL SKETCH	iv
ACKNOWLEDGEMENTS	v
CHAPTER 1. INTRODUCTION	1
CHAPTER 2. THE <i>DROSOPHILA</i> GUT MICROBIOTA PROVISIONS THIAMIN TO ITS HOST	22
CHAPTER 2.1 ABSTRACT	22
CHAPTER 2.2 IMPORTANCE	23
CHAPTER 2.3 INTRODUCTION	23
CHAPTER 2.4 RESULTS	27
CHAPTER 2.5 DISCUSSION	37
CHAPTER 2.6 MATERIALS AND METHODS.....	42
CHAPTER 2.7 ACKNOWLEDGMENTS	48
CHAPTER 2.8 REFERENCES	49
CHAPTER 2.9 SUPPLEMENTARY MATERIAL.....	52
CHAPTER 3. GENOMIC INSIGHTS INTO THE THIAMIN METABOLISM OF <i>PAENIBACILLUS THIAMINOLYTICUS</i> NRRL B-4156 and <i>P. APIARIUS</i> NRRL B-23460.....	62
CHAPTER 3.1 ABSTRACT	62
CHAPTER 3.2 ABBREVIATIONS	62
CHAPTER 3.3 INTRODUCTION	63
CHAPTER 3.4 ORGANISM INFORMATION	65
CHAPTER 3.5 GENOME SEQUENCE INFORMATION.....	71
CHAPTER 3.6 GENOME PROPERTIES.....	73
CHAPTER 3.7 INSIGHTS FROM THE GENOME	75
CHAPTER 3.8 CONCLUSIONS	82
CHAPTER 3.9 REFERENCES	83
CHAPTER 4. A ROLE FOR THIAMINASE I IN THIAMIN SALVAGE.....	88
CHAPTER 4.1 ABSTRACT	88
CHAPTER 4.2 IMPORTANCE	89
CHAPTER 4.3 INTRODUCTION	89
CHAPTER 4.4 RESULTS	94

CHAPTER 4.5 DISCUSSION.....	105
CHAPTER 4.6 MATERIALS AND METHODS.....	110
CHAPTER 4.7 REFERENCES	116
CHAPTER 4.8 SUPPLEMENTARY MATERIAL.....	120
CHAPTER 5. INVESTIGATIONS INTO THE METABOLIC AND BIOSYNTHETIC POTENTIAL OF ‘<i>CANDIDATUS EPULOPISCUM VIVIPAROUS</i>’	125
CHAPTER 5.1 ABSTRACT	125
CHAPTER 5.2 INTRODUCTION	126
CHAPTER 5.3 MATERIALS AND METHODS.....	130
CHAPTER 5.4 RESULTS	135
CHAPTER 5.5 DISCUSSION	191
CHAPTER 5.6 REFERENCES	198
APPENDIX A. INVESTIGATIONS INTO THE ROLE OF THIAMINASE I AS A PATHOGENICITY AND COMPETITION FACTOR	212
CHAPTER A.1 INTRODUCTION	212
CHAPTER A.2 MATERIALS AND METHODS.....	218
CHAPTER A.3 RESULTS	226
CHAPTER A.4 DISCUSSION	241
CHAPTER A.5 REFERENCES	249
APPENDIX B. ATTEMPTS AT CULTURING <i>EPULOPISCUM</i> SPP. AND THEIR RELATIVES	254
CHAPTER B.1 INTRODUCTION.....	254
CHAPTER B.2 CULTURING ATTEMPTS AT CORNELL UNIVERSITY	256
CHAPTER B.3 MATERIALS AND METHODS	256
CHAPTER B.4 RESULTS AND CONCLUSIONS.....	266
CHAPTER B.5 FIRST CULTURING ATTEMPTS AT LIZARD ISLAND.....	283
CHAPTER B.6 RESULTS AND CONCLUSIONS.....	283
CHAPTER B.7 CULTURING ATTEMPTS FROM MULTIPLE SURGEONFISH AT LIZARD ISLAND	286
CHAPTER B.8 MATERIALS AND METHODS	286
CHAPTER B.9 RESULTS AND CONCLUSIONS.....	287

CHAPTER B.10 DISCUSSION	309
CHAPTER B.11 REFERENCES.....	314

CHAPTER 1

INTRODUCTION

The profound impact bacteria have on the evolution of metazoans has only recently been appreciated and explored, and our understanding of how these two groups of organisms interact with and co-evolve is still developing (1, 2). It has been argued that the acquisition of the mitochondria provided cells with the energetics to expand their genomes and in turn invest energy into evolving new genes, regulatory systems, and protein families, all of which contributed to the ability to generate complex-multicellular life (3). The first animals evolved from protists around 635-800 million years ago, and this evolution occurred in a bacterial dominated world (1, 4), however, the potential bacterial influence on this development was not originally considered (5). Recent evidence suggests that bacteria may have played another role in the development of multicellularity and the evolution of animals as *Algoriphagus machipongonensis*, a member of the Bacteroidetes, and a prey bacterium to the choanoflagellate *Salpingoeca rosetta*, produces sulfonolipids that cause the choanoflagellate to initiate the formation of multicellular colonies (6). These colonies form not through aggregation, rather the development resembles that of an animal zygote as it occurs from a single founder cell. These sulfonolipids further stabilize this development and allow it to mature as a multicellular organism (7). Choanoflagellates are the closest living relative to animals, suggesting that environmental bacterial cues may have factored in on the development of multicellularity in animals (4, 7).

Bacteria influence multiple aspects of animal physiology. They produce developmental cues for some animals, exemplified by *Pseudoalteromonas luteoviolacea* inducing metamorphosis in the marine tubeworm (8) through the production of phage tail-like bacteriocins (9). In the parasitic nematode *Brugia malayi*, *Wolbachia* is required for the proper formation of anterior-

posterior polarity during zygote development, as *Wolbachia*-depleted cells have developmental division defects (10). Bacteria drive the development of organs in animals, epitomized by the morphogenesis of the light organ in the Hawaiian bobtail squid being induced and subsequently colonized by the bioluminescent bacterium *Vibrio fischeri* (11-13). Another example occurs in the parasitoid wasp, as *Wolbachia* transitioned from a facultative to an obligate symbiont because it is required for proper oocyte development within the ovaries (14). In mouse models, it has been demonstrated that a normal microbiota is necessary for complete gut development (15), and *Bacteroides fragilis* produces polysaccharide A while it colonizes the gut. This leads to proper maturation of the developing immune system (2, 16). In mammals, the microbiota has been implicated in proper neural development as the gut microbiota produces many neurotransmitters, and germ-free mice have altered neurotransmitter circulation and neurotransmitter receptor expression (15). The microbiota influences complex traits such as host mood, stress levels, and behavior (17). Germ-free mice display differences in behavior, such as riskier behavior in comparison to mice with a normal microbiota (18). In other studies, mice born from a maternal infection model display behavioral signs similar to autism spectrum disorder, and these signs are reduced upon colonization by *B. fragilis* (19).

Bacteria allow for animals to exploit niches they would not be able to exploit through nutritional symbioses. Bacterial members of the ruminal microbiota digest the indigestible components of their host diet (20), and supply the ruminants with short chain fatty acids (SCFAs), which can cover up to 80% of their daily energy requirement (21). *Bathymodiolus puteoserpentis* mussels found in deep sea ridges harbor symbionts that couple hydrogen oxidation to carbon fixation, providing their host with carbon (22). Similar associations are found in other hydrothermal vent dwelling animals including bivalves which couple the oxidation of reduced

sulfur compounds to carbon fixation (23) or the Yeti crab, which farms lithotrophic symbionts on its claws, that it then feeds upon (24). Symbionts allow phloem and sap feeding insects to exploit a nutrient poor diet by providing their host with amino acids (25). This is demonstrated in the pea aphid-*Buchnera* system, in which the *Buchnera* bacteria are harbored in a special organ within the aphid and provide essential amino acids to its host in a diet dependent manner, or when most of these amino acids are present in the diet, *Buchnera* production of amino acids is reduced (26). Human populations that consume a diet high in seaweed contain a species of Bacteroidetes *Bacteroides plebius*, which horizontally acquired the ability to degrade the complex polysaccharide porphyran found in red algae, allowing these individuals to potentially get energy from this recalcitrant polysaccharide through fermentation products generated from its degradation (27).

To truly understand the nature of these interactions between animals and bacteria, model systems need to be utilized. *Drosophila melanogaster* provides one of the most rigorous and robust systems for interrogating the different aspects that shape these interactions, and the nature of these interactions. There are a wealth of tools to understand the host perspective in this interaction. The *D. melanogaster* immune response and the pathways that control this are well defined (28-30). There are multiple highly effective genetic tools such as RNA interference for making gene knockdowns (31) and binary expression systems (32) that allow for researchers to understand the genetic pathways and underpinnings that shape host physiology. Another strong genetic resource is the *D. melanogaster* Genetic Reference Panel, which is composed of 205 genome-sequenced inbred lines with polymorphisms (33). This allows for genome wide association studies to be conducted. These tools not only allow for the enhancement of *Drosophila* genetics and physiology, but they also are strong resources for understanding how interactions with microbes affects

Drosophila genotype and phenotype, and the reciprocal interaction. So much of our understanding of how microbes and animals interact with one another is interfaced at the nutrition of both organisms, it is necessary to be able to completely control for dietary factors that can potentially shape these interactions. Recently, a robust, completely chemically defined holidic diet was developed for *D. melanogaster* (34), allowing for investigations into how each dietary component affects the host, the microbiota, and their interaction. *D. melanogaster* also is very amenable to manipulation of the microbiota as completely germ-free flies can be developed and maintained for many generations (35), as the microbiota is not essential to host survival, and flies can be re-associated with specific members of the microbiota rather easily (36). All these characteristics make *D. melanogaster* an ideal host candidate for studying host-bacteria interactions, and in particular the interaction of host, microbiota, and nutrition.

The microbiota of *D. melanogaster* facilitates this study as well. Unlike mammalian systems, the *D. melanogaster* community structure is much simpler and less diverse (37), and this diversity is even more diminished in lab-reared flies. Further the dominant members of the *D. melanogaster* microbiota are cultivable, and many have sequenced genomes (38). Although variable, lab reared *D. melanogaster* typically have microbiota communities dominated by Acetobacteraceae and Lactobacillaceae (37, 39-47). As flies age, total bacterial abundance increases (39) and the community composition shifts, as in one study, flies started out with a higher abundance in *Lactobacillus* species, but as the flies progressed, the community shifted to having *Acetobacter pomorum* being the most abundant community member (40). The relationship is not exactly stable as it appears that flies seed their food with their microbiota, which serves as a microbe reservoir. This reservoir allows for *D. melanogaster* to maintain and replenish its microbiota (47). When *D. melanogaster* is transferred repeatedly to sterile food, their total

microbial abundance diminishes (47) These organisms profoundly affect their host (28) and can even influence host behaviors such as mating preferences as the presence *Lactobacillus plantarum* alters sex pheromones influence mating behavior (48).

There is a complex tradeoff with the microbiota and immunity. Community members protect flies from infections from some oral pathogens (47), induce epithelial cell turnover and intestinal stem cell proliferation in the gut (39, 49), prime the immune system (50), stimulate antiviral immunity (51), and shape gut morphology (49). Despite these benefits, there is a negative aspect to this relationship, as *D. melanogaster* has to maintain its microbiota without compromising its immunity. Microbiota members trigger reactive oxygen species and antimicrobial peptide (AMP) production (50) which is done to manage the microbiota (28). As flies age and their bacterial load increases, the activation of AMP expression increases, and chronic expression of the immune response negatively impacts host health (52). Germ-free *D. melanogaster* have reduced levels of the immune response (49, 53), and have increased lifespans (53, 54). This is because they do not have the same microbiota induced stress in their guts. The microbiota induced damage leads to increased stem cell proliferation and epithelial turnover, to the point where epithelial turnover results in dysplasia and the intestinal barrier is no longer functional, and loss of barrier function eventually leads to unchecked proliferation and death of the host (28, 49, 53-55).

The microbiota influences host metabolism and can achieve this through the manipulation of host signaling pathways as revealed through a genome-wide association study using the *D. melanogaster* genetic reference panel (56). Other studies show that *A. pomorum* modulates the nutrient sensing insulin/insulin-like growth factor signaling pathway through the production of acetic acid via its periplasmic pyrroloquinoline quinone-dependent alcohol dehydrogenase.

Stimulation of this pathway accelerates host development and increases host body size (42). In amino acid poor conditions, *L. plantarum* mediates the Target of Rapamycin signaling pathway which also hastens growth rate and development (43). *Acetobacter* species influence host tri-acyl-glyceride levels (36, 57), and this is achieved through interaction with other community members as co-colonization of the gut with a *Lactobacillus* species is required for this response. *A. tropicalis* further influences host tri-acyl-glyceride levels through diet modification. Through the oxidation of dietary sugars, *A. tropicalis* alters glucose levels and thus lipid content of its host (58, 59). The microbiota can effect food choice, as flies on a defined diet lacking essential amino acids seek out a high yeast diet to supplement this. However, when certain microbiota members are present in concert, like *A. pomorum* and *L. plantarum* or *L. brevis*, the flies no longer seek a yeast diet as the microbiota suppresses this, though it does not appear to be provisioning essential amino acids (60). It also appears that the *D. melanogaster* microbiota can contribute to host nutrition through the supplement of certain B-vitamins. The microbiota provides folate to their host, as bleach-treated eggs were unable to develop on a defined diet lacking folate, while non-treated eggs were (34). Other studies suggest *Acetobacter* species function to provision riboflavin to the host (45, 61).

The contribution of thiamin (vitamin B₁) by the microbiota has been suggested in *Drosophila* and evidence suggests it occurs in some other insect symbioses (62, 63). However, it has yet to be confirmed in *D. melanogaster* if this occurs, and there are not many functional studies clearly demonstrating thiamin provision to animal hosts from the microbiota. In humans, genomic evidence suggests that this occurs in the colon, however, it has not been functionally described (64-66). Animals are unable to synthesize thiamin so they must acquire it exogenously either from their diet or from their microbiota, and many species of bacteria are able to synthesize thiamin (67). Thiamin deficiency leads to disease and death in animals (68-71). In the tsetse fly, there is a

thiamin related bacterial-bacterial interaction occurring that also impacts host physiology. The obligate symbiont, *Wigglesworthia morsitans*, which can synthesize thiamin, upregulates expression of thiamin biosynthesis genes under low nutrient events, while the auxotrophic secondary symbiont *Sodalis glossinidius* upregulates expression of genes for thiamin uptake (72, 73). In the presence of thiamin monophosphate (TMP), expression of thiamin biosynthesis genes and *W. morsitans* abundance decreases within the host (73). Thiamin production by one bacterium likely fosters the growth of the other symbiont and the host (72, 73). Despite this example, how thiamin shapes the interactions between bacteria and animals is not well understood, and how it shapes interactions between bacteria is also quite undefined. The rumen microbiota is a clear example of how bacteria can both positively and negatively influence their host through thiamin. Ruminants acquire the majority of their daily thiamin from their microbiota as it is released when rumen microbes are lysed in the abomasum, and is then absorbed in the small intestine (74-76). Rumen microbes can also cause maladies in their hosts through biotic thiamin degradation, as degradation induced thiamin deficiency caused by *Clostridium sporogenes* can result in cerebrocortical necrosis in ruminants such as sheep (77).

Biotic thiamin degradation is caused by thiaminases, enzymes that degrade thiamin to its constituent pyrimidine and thiazole moieties through a nucleophilic attack of the pyrimidine moiety. This leads to a base exchange of the thiazole with a nucleophile (78). Two classes of thiaminases occur in nature, the more widely distributed, intracellular thiaminase II is found in multiple species of bacteria (79), fungi (80), and plants (81). This enzyme restores an intact pyrimidine moiety from thiamin that has base degraded due to environmental factors (82). In contrast, thiaminase I enzymes lack shared sequence similarity with thiaminase II and differ in many ways. While thiaminase II utilizes H₂O as a nucleophile (82), thiaminase I utilizes a variety

of organic nucleophiles to catalyze the base exchange reaction (83, 84), and the enzyme appears to have a common ancestor with the thiamin binding protein of *Escherichia coli* (85). Thiaminase I is not an intracellular enzyme, rather it is extracellular and secreted through the general secretory pathway (86, 87), and is only known to be produced by a small subset of bacteria and amoeba, some of which are human pathogens (78). Bacterial thiaminase I has been implicated in thiamin deficiencies in animals as the first known thiaminase I producing bacteria were isolated from the feces of patients suffering from thiamin deficiency (83). Thiaminase I may also be responsible for, or exacerbate some cases of infant botulism, caused by the thiaminase I producing pathogen *Clostridium botulinum*, as both thiamin deficiency and botulinum toxin result in paralysis (88). Thiaminase I producing bacteria have been isolated from the viscera of clupeid fish (89). When these fish are consumed by salmonids, the thiamin degrading activity leads to the development of Early Mortality Syndrome, a thiamin deficiency that leads to the death at the early life stages in the salmonid offspring (89). Known thiaminase I producing bacteria are not always found associated with these fish with high thiaminase I activity, as it is possible they produce it from another source, making the relationship not well defined (90). Thiamin deficiency occurs in other natural populations of birds (69), large mammals (70), alligators, sea lions, and dolphins (78). The role of bacteria, and bacterial production of thiaminase I has not been investigated in these animals, and the biological function of thiaminase I is a mystery. Bacterial thiaminase I has yet to be definitively demonstrated to be a pathogenicity factor during infection in animals.

Thiamin shapes ecological interactions between bacteria. As already described in the tsetse fly model where a positive interaction occurs as the promotion of growth of one symbiont, in a thiamin related manner, affects the fitness of the other obligate symbiont, because together both symbionts increase host fitness (72, 73). The concentration of thiamin in many environments is

not well known, but it is believed to be rather scarce (78). Thiamin is sensitive to many abiotic factors including pH, heat, UV-light, inorganic bases, and heavy metals (91). This could further increase the scarcity of thiamin found in natural environments and result in both relationships of cooperation between organisms, or competition. It has also caused bacteria to evolve ways to salvage thiamin and utilize precursors for growth. *Pelagibacter ubique* HTCC1062 of the SAR11 clade is a natural thiamin auxotroph as it lacks the ability to make the pyrimidine moiety, hydroxymethyl pyrimidine (HMP) (92). It is able to offset this auxotrophy through the acquisition of HMP from the environment, which it prefers to grow on as opposed to thiamin. HMP is exuded by cyanobacteria, supplementing this bacteria with the necessary component to quench its thiamin auxotrophy (92). In other marine systems, bacteria are able to support the growth of picoeukaryotic phytoplankton through the production of precursors for synthesizing the thiazole moiety (93). These are examples of positive interactions fostered by bacterial production of precursors of thiamin, however it is unclear how soil bacteria interact in regards to thiamin.

Competition for thiamin has not been well studied in bacteria. It has been suggested that thiaminase I could serve as a competition factor as it would deplete environmental thiamin and potentially limit the growth of competitors in an environment where thiamin is likely already scarce. However, there is no empirical evidence to demonstrate this (78). Evidence supports that ecological competition for this nutrient occurs between bacterial species is the production of bacimethrin by some bacteria (94). Bacimethrin is an analog of HMP with 2' methoxy group. When it is combined with the thiazole moiety, it forms the antivitamin 2'-methoxythiamin which competes with essential enzymes that bind thiamin pyrophosphate, the active cofactor of thiamin, rendering them non-functional (95). Further, thiaminase I is found in the biosynthetic operon for bacimethrin, though its function is unknown. Biotic and abiotic derived toxic thiamin analogs have

led to bacteria evolving ways to pre-empt and repair these damaged and dangerous compounds (82, 96), though our understanding of this is still not well known. The role that thiaminase I plays in competition and environmental interactions with other bacteria still needs to be defined. Further it is unclear how this enzyme impacts thiamin metabolism. Studies within an ecological context need to be conducted to elucidate the biological role of this enzyme.

Fish are another attractive, yet understudied model for examining the ecological interactions of bacteria and animals. Unlike *D. melanogaster*, fish harbor complex microbiotas, but similar patterns have been observed in both types of organisms. Like in *D. melanogaster*, in zebrafish the microbiota stimulates the epithelial cells of the gastrointestinal tract, resulting in differentiation, maturation, and proliferation of these cells (97-99). The microbiota also helps prime and adapt the immune system (100), as axenic fish have a weaker immune response (97, 101). The microbiota affects nutrition by providing B-vitamins like cobalamin to the host (102), stimulating nutrient uptake and absorption (97, 98, 103), and producing digestive enzymes that aid in the digestion of dietary materials (104, 105). The microbiota can even help their hosts adapt to temperature changes through the production of proteinases (106). The interaction between microbiota and host is shaped by many host factors including age, sex, genotype, and phylogeny, however, a large driver of this is diet. The interaction of these host factors with the environmental factor of diet drive and shape the complexity of the community and the type of enzymes produced by the community (107).

There is a lack of functional analysis studying how the microbiota affects marine tropical herbivorous and omnivorous fish, and in particular, surgeonfish of the family Acanthuridae (107). It has long been considered that the microbiota contributes to the nutrition of these fish through the hindgut fermentation of complex polysaccharides found in their diet and supplying the host

with assimilable SCFAs (108). These fish eat diets mainly composed of algae (109), which contain a wealth of recalcitrant polysaccharides including carrageenan, agar, porphyran, laminarin, cellulose, and mannan, as well as low molecular weight compounds that are not assimilable like D-mannitol (110). Herbivorous and omnivorous surgeonfish like *Naso tonganus* have high concentrations of short chain fatty acids like acetate in their gut, and this is likely due to their microbiota (111). The microbiotas of surgeonfish are dominated by Firmicutes but also contain Proteobacteria and Bacteroidetes, and community composition appears to be constructed in a diet dependent manner (112).

The Firmicutes found in these fish tend to be of *Epulopiscium* spp. group (112-116), which are members of the Lachnospiraceae, falling in the XIVb cluster of the Clostridiales (117). The first *Epulopiscium* species, *Epulopiscium fishelsoni* was discovered in the guts of *Acanthurus nigrofusus* in the Red Sea, and were originally thought to be protists due to their unique shape and size (113). These cigar shaped cells can reach lengths of 600 μm and widths of 80 μm (116). More of these types of organisms were found in other herbivorous and omnivorous surgeonfish (115, 118), and were grouped into morphotypes based on size, shape, and reproduction strategy. These organisms were then determined to be bacteria and were placed into the Lachnospiraceae (117). It is unclear their functional role in the symbiosis with their host fish, but it is suspected that it may be nutritional. Generally, multiple species are found within the same fish species (115), though in some cases they are found in different regions of the gut (118). One exception to this is morphotype B, which is found only in *N. tonganus*, and is the only *Epulopiscium* species found in the guts of adult *N. tonganus* fish. Their association with the fish seems to be obligate for them, as none of these bacteria have been detected outside their hosts. However, the association is facultative for their host as starved fish lose these symbionts but can survive without them. This

association does appear to be important as the same species of surgeonfish from distinct geographical locations harbor the same *Epulopiscium* spp (118). These organisms are quite understudied and need to be investigated in an ecological context to understand their relationship with their host.

Despite our lack of knowledge about the nature of their symbiosis with surgeonfish, we have an understanding of their unique cell biology. They are a model system for studying cell biology at the extreme spectrum of the bacterial world. Despite there being a range of shapes (119) and sizes in bacteria, they are generally small and maintain a favorable surface area-to volume ratio, to overcome the diffusion dependent constraints on metabolic fluxes and protein-substrate interactions, and to compete for nutrients at very low concentrations (120). *Epulopiscium* spp. defy this and have certain adaptations to make their surface area to volume ratio more favorable, as the largest morphotypes can have volumes exceeding $2,000,000 \mu\text{m}^3$. One such adaptation is extremely invaginated membranes (116, 121, 122), which effectively increases surface area. *Epulopiscium* spp. also contain a massive amount of DNA organized as nucleoids located at the periphery of the cell (121, 122). Type B is the most well studied of the morphotypes and it is estimated that the largest of these cells, which can reach lengths of $300 \mu\text{m}$ and widths of $60 \mu\text{m}$, contain 250 pg of DNA and are extremely polyploid, containing 10's of thousands of copies of their chromosome (123). This is likely a consequence of their large size and may enable the cells to have quicker, local responses to external stimuli. *Epulopiscium* spp. also employ a diversity of shapes and sizes. Many morphotypes are cigar shaped ranging in widths, as some can have a more narrow appearance (type A1, type A2, and type C), while type B cells are more ellipsoid shaped, and type J cells are long, laterally compressed, and filamentous shaped (115, 118). Lateral compression may be an adaptive trait to increase surface-area to volume in these cells.

Another unique and distinguishing factor of *Epulopiscium* spp. is their lifecycle and mode of reproduction. Multiple *Epulopiscium* spp. rely solely on the production of multiple intracellular offspring (118, 122), and this is reminiscent of the mode of propagation of *Metabacterium polyspora*, which relies solely on the production of multiple endospores for reproduction (124). Typically *Epulopiscium* spp. produce 2 intracellular offspring, however, type B can produce as many as 12 intracellular daughter cells. Intracellular offspring formation shares multiple hallmarks with the endosporulation program employed by Firmicutes (125). Offspring development in type B mother-cells begins with DNA coalescing at the poles, where FtsZ then localizes and asymmetric bipolar division occurs, initiating the onset of offspring formation (Stage I) (122). Upon completion of septation (Stage II) in which only about 0.2% of polar DNA is captured by the new daughter cells, the rest of the mother-cell polar DNA equaling approximately 1% of the total mother cell DNA, is then translocated into each daughter and the septa take on a curved shape (Stage II-III) (122). The daughters are fully engulfed by the mother cell (Stage III), where they remain close to the poles of the mother cell and begin to develop (122). As the daughter cells develop, they migrate from the poles, increase in size, taking a 2:1 length to width ratio (Stage IV*), replicate their genome, and eventually fill the mother cell's cytoplasm (122, 126). The mother cell deteriorates and the daughter cells are released, and the cycle continues in a diurnal fashion, where the population appears synchronized (122). The genomic basis for this in type B has also been outlined, confirming that these process is derived from endosporulation (126). Reproduction of both type A1 and A2 occur in a diurnal cycle through the production of these offspring (114). Type C and J also reproduce in a diurnal fashion, however, unlike the A1, A2, and B morphotypes, they produce dormant offspring, which have all the hallmarks of true endospores (118). This complicated life cycle apparently synchronizes with host-feeding; as the host is active

so are their symbionts, however during the night, while the host is no longer active, the mother cells senesce as their daughter cells mature (118). Upon morning when the fish begins feeding again the daughter cells release, and in the case of the endospore formers they germinate into vegetative cells and perpetuate the cycle. However, some morphotypes like type J are still able to reproduce through binary fission. Due to their extreme qualities *Epulopiscium* provide the ability to serve as a model to examine questions about bacterial cell biology, in particular, how a fermentative heterotroph can reach such large sizes, as well as fundamental questions regarding bacterial cell functional compartmentalization and chromosome organization. Inferences on host metabolism and biosynthetic capabilities are still lacking as well. Understanding these questions in the context of its fish host may provide insight into how *Epulopiscium* spp. affect their hosts, and in turn how their hosts shape their interaction.

References

1. McFall-Ngai M, Hadfield MG, Bosch TC, Carey HV, Domazet-Lošo T, Douglas AE, Dubilier N, Eberl G, Fukami T, Gilbert SF. 2013. Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci U S A* 110:3229-3236.
2. Rosenberg E, Zilber-Rosenberg I. 2016. Microbes drive evolution of animals and plants: the hologenome concept. *MBio* 7:e01395-15.
3. Lane N, Martin W. 2010. The energetics of genome complexity. *Nature* 467:929-934.
4. Alegado RA, King N. 2014. Bacterial influences on animal origins. *Cold Spring Harbor perspectives in biology* 6:a016162.
5. Rokas A. 2008. The origins of multicellularity and the early history of the genetic toolkit for animal development. *Annual review of genetics* 42:235-251.
6. Alegado RA, Brown LW, Cao S, Dermejian RK, Zuzow R, Fairclough SR, Clardy J, King N. 2012. A bacterial sulfonolipid triggers multicellular development in the closest living relatives of animals. *elife* 1:e00013.
7. Woznica A, Cantley AM, Beemelmans C, Freinkman E, Clardy J, King N. 2016. Bacterial lipids activate, synergize, and inhibit a developmental switch in choanoflagellates. *Proceedings of the National Academy of Sciences*:201605015.
8. Huang Y, Callahan S, Hadfield MG. 2012. Recruitment in the sea: bacterial genes required for inducing larval settlement in a polychaete worm. *Scientific reports* 2:228.
9. Shikuma NJ, Pilhofer M, Weiss GL, Hadfield MG, Jensen GJ, Newman DK. 2014. Marine tubeworm metamorphosis induced by arrays of bacterial phage tail-like structures. *Science* 343:529-533.

10. Landmann F, Foster JM, Michalski ML, Slatko BE, Sullivan W. 2014. Co-evolution between an endosymbiont and its nematode host: *Wolbachia* asymmetric posterior localization and AP polarity establishment. *PLoS neglected tropical diseases* 8:e3096.
11. McFall-Ngai MJ, Ruby EG. 1991. Symbiont recognition and subsequent morphogenesis as early events in an animal-bacterial mutualism. *Science* 254:1491.
12. Altura MA, Heath-Heckman EA, Gillette A, Kremer N, Krachler AM, Brennan C, Ruby EG, Orth K, McFall-Ngai MJ. 2013. The first engagement of partners in the *Euprymna scolopes*–*Vibrio fischeri* symbiosis is a two-step process initiated by a few environmental symbiont cells. *Environmental microbiology* 15:2937-2950.
13. Kremer N, Philipp EE, Carpentier M-C, Brennan CA, Kraemer L, Altura MA, Augustin R, Häsler R, Heath-Heckman EA, Peyer SM. 2013. Initial symbiont contact orchestrates host-organ-wide transcriptional changes that prime tissue colonization. *Cell host & microbe* 14:183-194.
14. Dedeine F, Vavre F, Fleury F, Loppin B, Hochberg ME, Boulétreau M. 2001. Removing symbiotic *Wolbachia* bacteria specifically inhibits oogenesis in a parasitic wasp. *Proceedings of the National Academy of Sciences* 98:6247-6252.
15. Gilbert SF, Bosch TC, Ledón-Rettig C. 2015. Eco-Evo-Devo: developmental symbiosis and developmental plasticity as evolutionary agents. *Nature Reviews Genetics* 16:611.
16. Mazmanian SK, Liu CH, Tzianabos AO, Kasper DL. 2005. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* 122:107-118.
17. Cryan JF, Dinan TG. 2012. Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. *Nature reviews neuroscience* 13:701-712.
18. Heijtz RD, Wang S, Anuar F, Qian Y, Björkholm B, Samuelsson A, Hibberd ML, Forssberg H, Pettersson S. 2011. Normal gut microbiota modulates brain development and behavior. *Proceedings of the National Academy of Sciences* 108:3047-3052.
19. Hsiao EY, McBride SW, Hsien S, Sharon G, Hyde ER, McCue T, Codelli JA, Chow J, Reisman SE, Petrosino JF. 2013. Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. *Cell* 155:1451-1463.
20. Jewell KA, McCormick CA, Odt CL, Weimer PJ, Suen G. 2015. Ruminal bacterial community composition in dairy cows is dynamic over the course of two lactations and correlates with feed efficiency. *Applied and environmental microbiology* 81:4697-4710.
21. Gäbel G, Sehested J. 1997. SCFA transport in the forestomach of ruminants. *Comparative Biochemistry and Physiology Part A: Physiology* 118:367-374.
22. Petersen JM, Zielinski FU, Pape T, Seifert R, Moraru C, Amann R, Hourdez S, Girguis PR, Wankel SD, Barbe V. 2011. Hydrogen is an energy source for hydrothermal vent symbioses. *Nature* 476:176.
23. Roeselers G, Newton IL. 2012. On the evolutionary ecology of symbioses between chemosynthetic bacteria and bivalves. *Applied microbiology and biotechnology* 94:1-10.
24. Thurber AR, Jones WJ, Schnabel K. 2011. Dancing for food in the deep sea: bacterial farming by a new species of yeti crab. *PLoS One* 6:e26243.
25. Bennett GM, Moran NA. 2013. Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a phloem-feeding insect. *Genome biology and evolution* 5:1675-1688.

26. Russell CW, Poliakov A, Haribal M, Jander G, van Wijk KJ, Douglas AE. 2014. Matching the supply of bacterial nutrients to the nutritional demand of the animal host. *Proceedings of the Royal Society of London B: Biological Sciences* 281:20141163.
27. Hehemann J-H, Correc G, Barbeyron T, Helbert W, Czjzek M, Michel G. 2010. Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* 464:908-912.
28. Bonfini A, Liu X, Buchon N. 2016. From pathogens to microbiota: How *Drosophila* intestinal stem cells react to gut microbes. *Developmental & Comparative Immunology* 64:22-38.
29. Hoffmann JA. 2003. The immune response of *Drosophila*. *Nature* 426:33.
30. Lemaitre B, Hoffmann J. 2007. The host defense of *Drosophila melanogaster*. *Annu Rev Immunol* 25:697-743.
31. Clemens JC, Worby CA, Simonson-Leff N, Muda M, Maehama T, Hemmings BA, Dixon JE. 2000. Use of double-stranded RNA interference in *Drosophila* cell lines to dissect signal transduction pathways. *Proceedings of the National Academy of Sciences* 97:6499-6503.
32. del Valle Rodríguez A, Didiano D, Desplan C. 2012. Power tools for gene expression and clonal analysis in *Drosophila*. *Nature methods* 9:47-55.
33. Huang W, Massouras A, Inoue Y, Peiffer J, Ràmia M, Tarone AM, Turlapati L, Zichner T, Zhu D, Lyman RF. 2014. Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines. *Genome research* 24:1193-1208.
34. Piper MD, Blanc E, Leitão-Gonçalves R, Yang M, He X, Linford NJ, Hoddinott MP, Hopfen C, Soultoukis GA, Niemeyer C. 2014. A holidic medium for *Drosophila melanogaster*. *Nat Methods* 11:100-105.
35. Bakula M. 1969. The persistence of a microbial flora during postembryogenesis of *Drosophila melanogaster*. *J Invertebr Pathol* 14:365-374.
36. Newell PD, Douglas AE. 2014. Interspecies interactions determine the impact of the gut microbiota on nutrient allocation in *Drosophila melanogaster*. *Appl Environ Microbiol* 80:788-796.
37. Chandler JA, Lang JM, Bhatnagar S, Eisen JA, Kopp A. 2011. Bacterial communities of diverse *Drosophila* species: ecological context of a host-microbe model system. *PLoS genetics* 7:e1002272.
38. Newell PD, Chaston JM, Wang Y, Winans NJ, Sannino DR, Wong AC, Dobson AJ, Kagle J, Douglas AE. 2014. In vivo function and comparative genomic analyses of the *Drosophila* gut microbiota identify candidate symbiosis factors. *Front Microbiol* 5:576.
39. Broderick NA, Buchon N, Lemaitre B. 2014. Microbiota-induced changes in *Drosophila melanogaster* host gene expression and gut morphology. *MBio* 5:e01117-14.
40. Wong CNA, Ng P, Douglas AE. 2011. Low-diversity bacterial community in the gut of the fruitfly *Drosophila melanogaster*. *Environ Microbiol* 13:1889-1900.
41. Wong AC, Chaston JM, Douglas AE. 2013. The inconstant gut microbiota of *Drosophila* species revealed by 16S rRNA gene analysis. *ISME J* 7:1922-1932.
42. Shin SC, Kim S-H, You H, Kim B, Kim AC, Lee K-A, Yoon J-H, Ryu J-H, Lee W-J. 2011. *Drosophila* microbiome modulates host developmental and metabolic homeostasis via insulin signaling. *Science* 334:670-674.

43. Storelli G, Defaye A, Erkosar B, Hols P, Royet J, Leulier F. 2011. *Lactobacillus plantarum* promotes *Drosophila* systemic growth by modulating hormonal signals through TOR-dependent nutrient sensing. *Cell Metab* 14:403-414.
44. Ren C, Webster P, Finkel SE, Tower J. 2007. Increased internal and external bacterial load during *Drosophila* aging without life-span trade-off. *Cell Metab* 6:144-152.
45. Fridmann-Sirkis Y, Stern S, Elgart M, Galili M, Zeisel A, Shental N, Soen Y. 2014. Delayed development induced by toxicity to the host can be inherited by a bacterial-dependent, transgenerational effect. *Front Genet* 5:27.
46. Broderick NA, Lemaître B. 2012. Gut-associated microbes of *Drosophila melanogaster*. *Gut microbes* 3:307-321.
47. Blum JE, Fischer CN, Miles J, Handelsman J. 2013. Frequent replenishment sustains the beneficial microbiome of *Drosophila melanogaster*. *MBio* 4:e00860-13.
48. Sharon G, Segal D, Ringo JM, Hefetz A, Zilber-Rosenberg I, Rosenberg E. 2010. Commensal bacteria play a role in mating preference of *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* 107:20051-20056.
49. Buchon N, Broderick NA, Chakrabarti S, Lemaître B. 2009. Invasive and indigenous microbiota impact intestinal stem cell activity through multiple pathways in *Drosophila*. *Genes Dev* 23:2333-2344.
50. Ryu J-H, Kim S-H, Lee H-Y, Bai JY, Nam Y-D, Bae J-W, Lee DG, Shin SC, Ha E-M, Lee W-J. 2008. Innate immune homeostasis by the homeobox gene caudal and commensal-gut mutualism in *Drosophila*. *Science* 319:777-782.
51. Sansone CL, Cohen J, Yasunaga A, Xu J, Osborn G, Subramanian H, Gold B, Buchon N, Cherry S. 2015. Microbiota-dependent priming of antiviral intestinal immunity in *Drosophila*. *Cell host & microbe* 18:571-581.
52. Paredes JC, Welchman DP, Poidevin M, Lemaître B. 2011. Negative regulation by amidase PGRPs shapes the *Drosophila* antibacterial response and protects the fly from innocuous infection. *Immunity* 35:770-779.
53. Guo L, Karpac J, Tran SL, Jasper H. 2014. PGRP-SC2 promotes gut immune homeostasis to limit commensal dysbiosis and extend lifespan. *Cell* 156:109-122.
54. Clark RI, Salazar A, Yamada R, Fitz-Gibbon S, Morselli M, Alcaraz J, Rana A, Rera M, Pellegrini M, William WJ. 2015. Distinct shifts in microbiota composition during *Drosophila* aging impair intestinal function and drive mortality. *Cell Rep* 12:1656-1667.
55. Biteau B, Hochmuth CE, Jasper H. 2008. JNK activity in somatic stem cells causes loss of tissue homeostasis in the aging *Drosophila* gut. *Cell Stem Cell* 3:442-455.
56. Dobson AJ, Chaston JM, Newell PD, Donahue L, Hermann SL, Sannino DR, Westmiller S, Wong AC-N, Clark AG, Lazzaro BP. 2015. Host genetic determinants of microbiota-dependent nutrition revealed by genome-wide analysis of *Drosophila melanogaster*. *Nature communications* 6.
57. Chaston JM, Dobson AJ, Newell PD, Douglas AE. 2016. Host genetic control of the microbiota mediates the *Drosophila* nutritional phenotype. *Appl Environ Microbiol* 82:671-679.
58. Chaston JM, Newell PD, Douglas AE. 2014. Metagenome-wide association of microbial determinants of host phenotype in *Drosophila melanogaster*. *MBio* 5:e01631-14.
59. Huang J-H, Douglas AE. 2015. Consumption of dietary sugar by gut bacteria determines *Drosophila* lipid content. *Biol Lett* 11:20150469.

60. Leitão-Gonçalves R C-SZ, Francisco AP, Fioreze GT, Anjos M, Baltazar C, et al. 2017. Commensal bacteria and essential amino acids control food choice behavior and reproduction. *PLoS Biol* 15:e2000862.
61. Wong AC-N, Dobson AJ, Douglas AE. 2014. Gut microbiota dictates the metabolic response of *Drosophila* to diet. *J Exp Biol* 217:1894-1901.
62. Harington J. 1960. Synthesis of thiamine and folic acid by *Nocardia rhodnii*, the micro-symbiont of *Rhodnius prolixus*. *Nature* 188:1027-1028.
63. Salem H, Bauer E, Strauss AS, Vogel H, Marz M, Kaltenpoth M. 2014. Vitamin supplementation by gut symbionts ensures metabolic homeostasis in an insect host. *Proc R Soc Lond B Biol Sci* 281:20141838.
64. Nabokina SM, Inoue K, Subramanian VS, Valle JE, Yuasa H, Said HM. 2014. Molecular identification and functional characterization of the human colonic thiamine pyrophosphate transporter. *J Biol Chem* 289:4405-4416.
65. Magnúsdóttir S, Ravcheev D, de Crécy-Lagard V, Thiele I. 2015. Systematic genome assessment of B-vitamin biosynthesis suggests co-operation among gut microbes. *Front Genet* 6:148.
66. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto J-M. 2011. Enterotypes of the human gut microbiome. *Nature* 473:174-180.
67. Jurgenson CT, Begley TP, Ealick SE. 2009. The structural and biochemical foundations of thiamin biosynthesis. *Annu Rev Biochem* 78:569-603.
68. Edwin E, Jackman R. 1970. Thiaminase I in the development of cerebrocortical necrosis in sheep and cattle. *Nature* 228:772-774.
69. Balk L, Hägerroth P-Å, Åkerman G, Hanson M, Tjärnlund U, Hansson T, Hallgrímsson GT, Zebühr Y, Broman D, Mörner T. 2009. Wild birds of declining European species are dying from a thiamine deficiency syndrome. *Proc Natl Acad Sci U S A* 106:12001-12006.
70. Balk L, Hägerroth P-Å, Gustavsson H, Sigg L, Åkerman G, Muñoz YR, Honeyfield DC, Tjärnlund U, Oliveira K, Ström K. 2016. Widespread episodic thiamine deficiency in Northern Hemisphere wildlife. *Sci Rep* 6.
71. Evans C, Carlosn W, Green R. 1942. The Pathology of Chastek Paralysis in Foxes: A Counterpart of Wernicke's Hemorrhagic Pollioencephalitis of Man. *The American journal of pathology* 18:79.
72. Snyder AK, Deberry JW, Runyen-Janecky L, Rio RV. 2010. Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts. *Proc R Soc Lond B Biol Sci* 277:2389-2397.
73. Snyder AK, McLain C, Rio RV. 2012. The tsetse fly obligate mutualist *Wigglesworthia morsitans* alters gene expression and population density via exogenous nutrient provisioning. *Applied and environmental microbiology* 78:7792-7797.
74. Breves G, Brandt M, Hoeller H, Rohr K. 1981. Flow of thiamin to the duodenum in dairy cows fed different rations. *J Agric Sci* 96:587-591.
75. Breves G, Hoeller H, Harmeyer J, Martens H. 1979. Gastro-intestinal passage and balance of thiamin in sheep. *Ann Rech Vet* 10:465-466.
76. Breves G, Hoeller H, Harmeyer J, Martens H. 1980. Thiamin balance in the gastrointestinal tract of sheep. *J Anim Sci* 51:1177-1181.
77. Shreeve J, Edwin E. 1974. Thiaminase-producing strains of *Cl. Sporogenes* associated with outbreaks of cerebrocortical necrosis. *The Veterinary record* 94:330.

78. Kraft CE, Angert ER. 2017. Competition for vitamin B1 (thiamin) structures numerous ecological interactions. *The Quarterly Review of Biology* 92:151-168.
79. Toms AV, Haas AL, Park J-H, Begley TP, Ealick SE. 2005. Structural characterization of the regulatory proteins TenA and TenI from *Bacillus subtilis* and identification of TenA as a thiaminase II. *Biochemistry* 44:2319-2329.
80. Onozuka M, Konno H, Kawasaki Y, Akaji K, Nosaka K. 2007. Involvement of thiaminase II encoded by the THI20 gene in thiamin salvage of *Saccharomyces cerevisiae*. *FEMS yeast research* 8:266-275.
81. Zallot R, Yazdani M, Goyer A, Ziemak MJ, Guan J-C, McCarty DR, de Crécy-Lagard V, Gerdes S, Garrett TJ, Benach J. 2014. Salvage of the thiamin pyrimidine moiety by plant TenA proteins lacking an active-site cysteine. *Biochemical Journal* 463:145-155.
82. Jenkins AH, Schyns G, Potot S, Sun G, Begley TP. 2007. A new thiamin salvage pathway. *Nature chemical biology* 3:492-497.
83. Fujita A. 1954. Thiaminase. *Advances in Enzymology and Related Areas of Molecular Biology*, Volume 15:389-421.
84. Douthit H, Airth R. 1966. Thiaminase I of *Bacillus thiaminolyticus*. *Archives of biochemistry and biophysics* 113:331-337.
85. Soriano EV, Rajashankar KR, Hanes JW, Bale S, Begley TP, Ealick SE. 2008. Structural similarities between thiamin-binding protein and thiaminase-I suggest a common ancestor. *Biochemistry* 47:1346-1357.
86. Costello CA, Kelleher NL, Abe M, McLafferty FW, Begley TP. 1996. Mechanistic studies on thiaminase I overexpression and identification of the active site nucleophile. *Journal of Biological Chemistry* 271:3445-3452.
87. Agee CC, Airth R. 1973. Reversible inactivation of thiaminase I of *Bacillus thiaminolyticus* by its primary substrate, thiamine. *Journal of bacteriology* 115:957-965.
88. Ringe H, Schuelke M, Weber S, Dorner BG, Kirchner S, Dorner MB. 2014. Infant botulism: is there an association with thiamine deficiency? *Pediatrics* 134:e1436-e1440.
89. Honeyfield DC, Hinterkopf JP, Brown SB. 2002. Isolation of thiaminase-positive bacteria from alewife. *Transactions of the American Fisheries Society* 131:171-175.
90. Richter CA, Evans AN, Wright-Osment MK, Zajicek JL, Heppell SA, Riley SC, Krueger CC, Tillitt DE. 2012. *Paenibacillus thiaminolyticus* is not the cause of thiamine deficiency impeding lake trout (*Salvelinus namaycush*) recruitment in the Great Lakes. *Canadian Journal of Fisheries and Aquatic Sciences* 69:1056-1064.
91. Dwivedi BK, Arnold RG. 1973. Chemistry of thiamine degradation on food products and model systems. Review. *Journal of Agricultural and Food Chemistry* 21:54-60.
92. Carini P, Campbell EO, Morré J, Sanudo-Wilhelmy SA, Thrash JC, Bennett SE, Temperton B, Begley T, Giovannoni SJ. 2014. Discovery of a SAR11 growth requirement for thiamin's pyrimidine precursor and its distribution in the Sargasso Sea. *ISME J* 8:1727-1738.
93. Paerl RW, Bouget F-Y, Lozano J-C, Vergé V, Schatt P, Allen EE, Palenik B, Azam F. 2017. Use of plankton-derived vitamin B1 precursors, especially thiazole-related precursor, by key marine picoeukaryotic phytoplankton. *The ISME journal* 11:753.
94. Cooper LE, O'Leary SnE, Begley TP. 2014. Biosynthesis of a Thiamin Antivitamin in *Clostridium botulinum*. *Biochemistry* 53:2215-2217.

95. Reddick JJ, Saha S, Lee J-m, Melnick JS, Perkins J, Begley TP. 2001. The mechanism of action of bacimethrin, a naturally occurring thiamin antimetabolite. *Bioorganic & medicinal chemistry letters* 11:2245-2248.
96. Goyer A, Hasnain G, Frelin O, Ralat MA, Gregory JF, Hanson AD. 2013. A cross-kingdom Nudix enzyme that pre-empts damage in thiamin metabolism. *Biochemical Journal* 454:533-542.
97. Rawls JF, Samuel BS, Gordon JI. 2004. Gnotobiotic zebrafish reveal evolutionarily conserved responses to the gut microbiota. *Proceedings of the National Academy of Sciences of the United States of America* 101:4596-4601.
98. Bates JM, Mittge E, Kuhlman J, Baden KN, Cheesman SE, Guillemin K. 2006. Distinct signals from the microbiota promote different aspects of zebrafish gut differentiation. *Developmental biology* 297:374-386.
99. Cheesman SE, Neal JT, Mittge E, Seredick BM, Guillemin K. 2011. Epithelial cell proliferation in the developing zebrafish intestine is regulated by the Wnt pathway and microbial signaling via Myd88. *Proceedings of the National Academy of Sciences* 108:4570-4577.
100. Bates JM, Akerlund J, Mittge E, Guillemin K. 2007. Intestinal alkaline phosphatase detoxifies lipopolysaccharide and prevents inflammation in zebrafish in response to the gut microbiota. *Cell host & microbe* 2:371-382.
101. Pérez T, Balcázar J, Ruiz-Zarzuela I, Halaihel N, Vendrell D, De Blas I, Múzquiz J. 2010. Host-microbiota interactions within the fish intestinal ecosystem. *Mucosal immunology*.
102. Sugita H, Miyajima C, Deguchi Y. 1991. The vitamin B12-producing ability of the intestinal microflora of freshwater fish. *Aquaculture* 92:267-276.
103. Semova I, Carten JD, Stombaugh J, Mackey LC, Knight R, Farber SA, Rawls JF. 2012. Microbiota regulate intestinal absorption and metabolism of fatty acids in the zebrafish. *Cell host & microbe* 12:277-288.
104. Nayak SK. 2010. Role of gastrointestinal microbiota in fish. *Aquaculture Research* 41:1553-1573.
105. Wang AR, Ran C, Ringø E, Zhou ZG. 2017. Progress in fish gastrointestinal microbiota research. *Reviews in Aquaculture*.
106. Kuz'mina V, Pervushina K. The role of proteinases of the enteral microbiota in temperature adaptation of fish and helminths, p 326-328. *In* (ed), Springer,
107. Clements KD, Angert ER, Montgomery WL, Choat JH. 2014. Intestinal microbiota in fishes: what's known and what's not. *Molecular ecology* 23:1891-1898.
108. Mountfort DO, Campbell J, Clements KD. 2002. Hindgut fermentation in three species of marine herbivorous fish. *Applied and environmental microbiology* 68:1374-1380.
109. Choat J, Clements K, Robbins W. 2002. The trophic status of herbivorous fishes on coral reefs. *Marine Biology* 140:613-623.
110. Hehemann J-H, Boraston AB, Czjzek M. 2014. A sweet new wave: structures and mechanisms of enzymes that digest polysaccharides from marine algae. *Current opinion in structural biology* 28:77-86.
111. Clements K, Choat J. 1995. Fermentation in tropical marine herbivorous fishes. *Physiological Zoology* 68:355-378.
112. Miyake S, Ngugi DK, Stingl U. 2015. Diet strongly influences the gut microbiota of surgeonfishes. *Molecular ecology* 24:656-672.

113. Fishelson L, Montgomery WL, Myrberg Jr AA. 1985. A unique symbiosis in the gut of tropical herbivorous surgeonfish(Acanthuridae: Teleostei) from the Red Sea. *Science(Washington)* 229:49-51.
114. Montgomery W, POLLAK PE. 1988. *Epulopiscium fishelsoni* ng, n. sp., a protist of uncertain taxonomic affinities from the gut of an herbivorous reef fish. *Journal of Eukaryotic Microbiology* 35:565-569.
115. Clements K, Sutton D, Choat J. 1989. Occurrence and characteristics of unusual protistan symbionts from surgeonfishes (Acanthuridae) of the Great Barrier Reef, Australia. *Marine Biology* 102:403-412.
116. Clements K, Bullivant S. 1991. An unusual symbiont from the gut of surgeonfishes may be the largest known prokaryote. *Journal of bacteriology* 173:5359-5362.
117. Angert ER, Clements KD, Pace NR. 1993. The largest bacterium. *Nature* 362:239-241.
118. Flint JF, Drzymalski D, Montgomery WL, Southam G, Angert ER. 2005. Nocturnal production of endospores in natural populations of *Epulopiscium*-like surgeonfish symbionts. *Journal of bacteriology* 187:7460-7470.
119. Young KD. 2006. The selective value of bacterial shape. *Microbiology and molecular biology reviews* 70:660-703.
120. Schulz HN, Jørgensen BB. 2001. Big bacteria. *Annual Reviews in Microbiology* 55:105-137.
121. Robinow C, Angert ER. 1998. Nucleoids and coated vesicles of “*Epulopiscium*” spp. *Archives of microbiology* 170:227-235.
122. Angert ER, Clements KD. 2004. Initiation of intracellular offspring in *Epulopiscium*. *Molecular microbiology* 51:827-835.
123. Mendell JE, Clements KD, Choat JH, Angert ER. 2008. Extreme polyploidy in a large bacterium. *Proceedings of the National Academy of Sciences* 105:6730-6734.
124. Angert ER, Losick RM. 1998. Propagation by sporulation in the guinea pig symbiont *Metabacterium polyspora*. *Proceedings of the National Academy of Sciences* 95:10218-10223.
125. Tan IS, Ramamurthi KS. 2014. Spore formation in *Bacillus subtilis*. *Environmental microbiology reports* 6:212-225.
126. Miller DA, Suen G, Clements KD, Angert ER. 2012. The genomic basis for the evolution of a novel form of cellular reproduction in the bacterium *Epulopiscium*. *BMC genomics* 13:265.

CHAPTER 2

THE *DROSOPHILA MELANOGASTER* GUT MICROBIOTA PROVISIONS THIAMIN TO ITS HOST*

2.1 ABSTRACT

The microbiota of *Drosophila melanogaster* has a substantial impact on host physiology and nutrition. There is evidence that these effects are attributable to vitamin provision, but the relationships between microbe-derived vitamins and host health remain to be established systematically. We explored the contribution of microbiota to the interaction between host and the dietary nutrient thiamin (vitamin B₁) at different stages of the fly lifecycle. Using chemically defined diets varying in available thiamin, we found that the interaction of thiamin concentration and microbiota did not significantly affect the longevity of adult *D. melanogaster* flies. Further, the interplay between dietary thiamin and presence of the microbiota did not have an impact on egg production of young or older females. However, we determined that thiamin availability has a large impact on offspring development, as axenic offspring were unable to develop into pupae on a diet lacking thiamin. Flies on this diet were rescued when the microbiota was present or added back, demonstrating that the microbiota was able to provide enough thiamin to the host to support development. Through gnotobiotic studies, we determined that only *Acetobacter pomorum*, a common member of the microbiota, was able to rescue development of fly larvae raised on the no-thiamin diet. It was also the only microbiota member that produced measurable amounts of thiamin when grown on the thiamin-free fly medium. The results demonstrate that the *D. melanogaster* microbiota functions to provision thiamin to its host in a low-thiamin environment.

*submitted to mBio where it is currently in the review process, the chapter is written according to their guidelines

2.2 IMPORTANCE

There has been a long-standing assumption that the microbiota of animals provides their hosts with essential B-vitamins, however, there is not a wealth of empirical evidence supporting this idea, especially for vitamin B₁ (thiamin). To determine if this is the case, we used *Drosophila melanogaster* and chemically defined diets varying in thiamin concentration as a model. We found that the microbiota does provide thiamin to its host, enough so to allow for the development of flies on a thiamin-free diet. The power of the *Drosophila*-microbiota system allowed us to determine that one microbiota member in particular, *Acetobacter pomorum*, is responsible for the thiamin provisioning. Thereby, our study verifies this long-standing hypothesis. Finally, the methods used in this work are applicable for interrogating the underpinnings of other aspects of the tripartite interaction between diet, host, and microbiota.

2.3 INTRODUCTION

The interplay between animals and microbes has helped forge the evolution of metazoa (1). Throughout their life, animals are in constant contact with microorganisms and these interactions are dynamic. The microbiota influences immune, developmental and metabolic functions of the host, through a plethora of secondary metabolites and molecules, however it often remains unclear what secondary metabolites and what metabolic functions are affected by which member of the microbiota.

Recently, *Drosophila melanogaster* has emerged as a potent model to study the function of gut microbes (2). The gut microbiota of laboratory reared *D. melanogaster* is variable (3, 4), but this low-diversity community is generally dominated by a few members of the Lactobacillaceae and Acetobacteraceae (3, 5-9). The microbiota impacts host physiology and

development as it primes the immune system (10), shapes gut morphology and homeostasis (3, 10, 11), and even influences mating preferences (12). In *D. melanogaster*, the nutritional impacts of its microbiota are substantial (9, 13, 14). Through manipulation of host signaling pathways, *Acetobacter* spp. can accelerate host development, increase body size and growth rate, and help regulate host glucose and lipid levels (6, 9, 15, 16). *Acetobacter pomoroum* possibly achieves this through the production of acetic acid, which influences the nutrient sensing insulin/insulin-like growth factor signaling pathway (6). When *D. melanogaster* is raised in nutrient-poor conditions, *Lactobacillus plantarum* modulates the activity of the Target of Rapamycin (TOR) signaling pathway, enhancing the production of hormonal signals that hasten larval development and growth rate (7). Microbiota composition also influences host nutritional phenotypes (17). For instance, *Acetobacter* spp. maintain tri-acyl-glycerol (TAG) levels inside the host in a dose-dependent manner, and this metabolic response is stimulated by co-colonization of the gut with *Lactobacillus* species (16, 17). Microbiota driven dietary modification, particularly the reduction of dietary glucose by microbial oxidation, also accounts for metabolic responses of the host (18, 19).

Despite the positive impacts the microbiota has on *D. melanogaster* nutrition and physiology, there are tradeoffs imposed by this close association. In the gut, *Drosophila* produces reactive oxygen species (20) and antimicrobial peptides (AMPs) (21) to combat infection and control gut microbiota (10), however, the stress induced by microbial presence damages the epithelium, leading to increased epithelial cell turnover (11). As flies age, their bacterial load increases (3, 8, 20, 22), contributing to a shift from epithelium renewal to dysplasia, which eventually compromises gut integrity (11, 20, 22, 23). The onset of the loss of barrier function coincides with an increase in γ -Proteobacteria at the expense of Firmicutes numbers (22). Barrier dysfunction can in turn lead to increased α -Proteobacteria (*Acetobacter*) abundance, a concomitant

rise in the immune response, and ultimately host death (20, 22). Aging also results in a loss of immune regulation, with the persistent activation of the Immune Deficiency (Imd) pathway, resulting in chronic expression of AMPs (11, 20). In contrast, axenic flies have a depressed immune response (11, 20) and a longer lifespan than their conventionally raised counterparts (22). Old axenic flies exhibit healthier, less deteriorated guts as there is reduced epithelial dysplasia, and intestinal stem cell proliferation rates remain similar to rates seen in younger flies (11, 20).

It has long been hypothesized that the gut microbiota of animals provides B vitamins to their hosts (24). Recent studies have implicated the *Drosophila* microbiota in supplying folate (14, 25) and riboflavin (9, 13), however, the contribution of thiamin (B₁) from the microbiota is not well understood. Thiamin is necessary for cellular life, as its diphosphorylated form, thiamin pyrophosphate (TPP), serves as a cofactor for enzymes involved in essential cellular processes including energy metabolism and the biosynthesis of the precursors for nucleotides, amino acids, and other cellular compounds (26-28). Consequently, thiamin deficiency results in disease and death in animals, including humans (28-31). Many bacteria, archaea, fungi, and plants, can synthesize thiamin (32), while animals typically acquire thiamin through their diet. In ruminants, however, the majority of the thiamin entering the duodenum originates from the rumen microbiota (33-35); this thiamin is released when rumen microbes are lysed in the abomasum making it available for absorption in the small intestine. In humans, dietary thiamin is taken up by transporters in small intestine enterocytes (36) and in the colon where colonocytes have an uptake system specific for TPP (26). Genomic evidence indicates that colonic bacteria are enriched in enzymes for thiamin synthesis, suggesting that the gut microbiota may be a thiamin source for the host (37, 38). However, a comprehensive understanding of how the microbiota impact thiamin metabolism in animal systems is still lacking.

Drosophila melanogaster provides a rigorous genetic and physiological system for interrogating the relationship between an animal host, its microbiota, and diet. Their low diversity microbial community, the ease of generating axenic animals (39), and reconstituting microbiota in axenic flies allows for the functional characterization and identification of dietary contributions of individual members of the microbiota (3, 16, 17). With the advent of a robust chemically defined diet (25, 40), it is now possible to fully examine how each dietary component contributes to this intricate tripartite interaction (40).

In this study, we used a chemically defined diet with conventionally raised (CR) and axenic (Ax) flies to determine the impact of dietary thiamin and microbiota on the development, longevity and egg production of adult *D. melanogaster*. We found that Ax flies had a longer lifespan than their CR counterparts, but this was not thiamin-dependent. There was also no discernable effect on female egg production. The larval period greatly impacts adult fitness as larval stores are carried into adulthood (41), and so we investigated the impact of dietary thiamin and the microbiota on development of the progeny from eggs. Axenic larvae developing on the thiamin-free diet were unable to pupate, however, CR larvae were able to pupate and develop into adults. This demonstrated that the microbiota provisioned thiamin to its host, allowing for larval development and growth into adulthood. To determine which members of the microbiota provided thiamin, we generated mono-associated flies and followed development on the thiamin-free diet. We found that development was rescued only when eggs were inoculated with *A. pomorum*, either in mono- or poly-association, and that *A. pomorum* was the only microbiota member that produced thiamin on the chemically defined diet. These results establish that the *D. melanogaster* microbiota functions as a source of thiamin for its host in a low thiamin environment.

2.4 RESULTS

The interaction of thiamin and microbiota do not significantly affect adult lifespan

In animals, thiamin is essential for energy metabolism and proper cell function (28, 42, 43). We thus hypothesized that thiamin deficiency would alter the lifespan of *Drosophila*, but that microbial thiamin provision would rescue this effect. To determine the role of thiamin contents in the food, and the impact of gut microbes on thiamin availability, Ax and CR flies were fed one of four chemically defined diets, varying in thiamin (0.00, 0.04, 0.2, or 1.0 $\mu\text{g/mL}$). Consistent with previous reports, Ax flies on all diets were longer-lived than CR flies (Fig. 1), which may be attributed to deleterious effects of the microbiota on gut barrier function (11, 20, 22). When comparing Ax and CR flies on each diet, we observed significantly longer lifespans of the Ax flies on the 0.04, 0.2 and 1.0 $\mu\text{g/mL}$ thiamin diets, however, there was no statistically significant difference between the lifespans of the Ax and CR flies on the no-thiamin diet (Fig. 1). Increasing levels of dietary thiamin linearly decreased median lifespan of CR flies, suggesting a toxicity of thiamin in these conditions. In addition, the absence of microbiota altered the effect of dietary thiamin on lifespan ($p = 0.0002$), but there was no significant effect of the thiamin by microbiota interaction term. Median lifespan of Ax flies was increased by even the smallest addition of dietary thiamin (Fig. 1). Thus, an absence of dietary thiamin is limiting to Ax fly lifespan, but beneficial for lifespan under CR conditions. Since Ax flies were longer-lived in general, but rendered sensitive to dietary thiamin in the opposite direction to CR flies, our data indicate that the influence of the microbiota on longevity is determined by the intersection of immunological and nutritional factors.

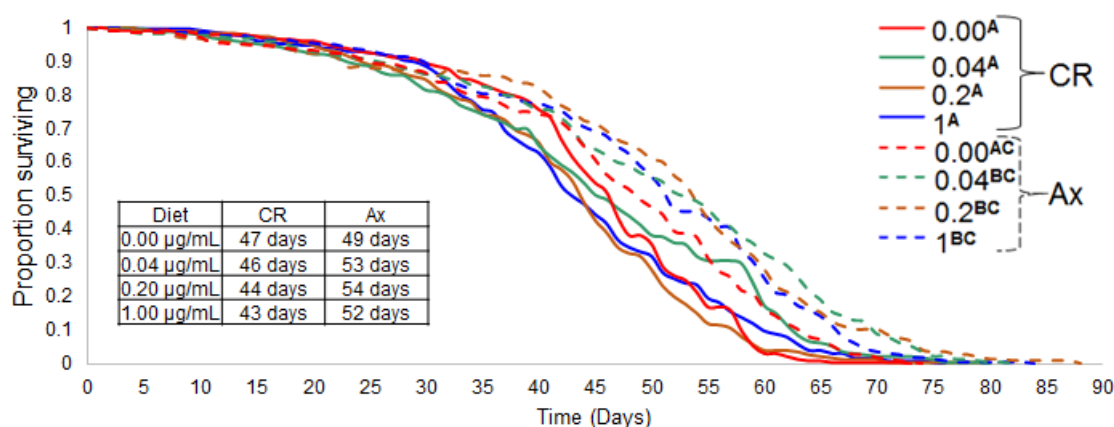


FIG 1 Impact of dietary thiamin on the longevity of conventionally reared (CR) and axenic (Ax) flies. Survival curves of CR and Ax flies on four diets ranging in vitamin B₁ concentration (0.00, 0.04, 0.2, and 1.0 µg/mL) indicate little difference between treatments. The 95% Wald Confidence Limits (CL) show significant differences when comparing CR and Ax flies to each other on the 0.04, 0.2, and 1.0 µg/mL thiamin diets. Treatments designated with the same superscript letters on the figure legends are not significantly different, for example those with the AC label are not significantly different than the treatments just with A or C superscripts. The insert table includes the median survival for the CR and Ax flies. Point estimates and CL are provided in Table S1.

Dietary thiamin and microbiota do not impact female egg production

In insects, including *D. melanogaster*, egg production is an energetically intensive process impacted by female metabolic status (44, 45), and promoted by dietary B vitamins (25, 46). Therefore, we hypothesized that egg production (measured by egg laying) could be conditional on the supply of dietary thiamin, and may be enhanced by the presence of the microbiota. Specifically, we predicted that Ax flies would lay fewer eggs than CR flies, and egg output would be disproportionately impacted by reduction of dietary thiamin in Ax flies relative to CR flies. To test this, we counted eggs laid by flies from the longevity study and analyzed egg production via an ANOVA (Fig. 2). For each diet, the average number of eggs produced by 4-day-old females were lower for Ax than CR flies ($p < 0.01$), but this was not diet-by-microbiota-dependent (Fig. 2A). These findings suggest that in early adulthood, there is no interaction between thiamin and

microbiota affecting egg production; rather it is the presence of the microbiota that promotes fecundity, independent of thiamin levels.

It has been previously shown that starved *D. melanogaster* females are able to produce eggs into the sixth day of adulthood from nutrients stored from their larval period (47). This raises the possibility that we did not detect an effect of a lack of dietary thiamin on egg production due to thiamin stores in the fly. As flies age and exhaust their larval supplies, they could become more reliant on their diet for nutrition and egg production. Based on this, we hypothesized that older, axenic females would produce fewer eggs when dietary thiamin was lacking, but this defect could be alleviated by the presence of the microbiota. Therefore, we examined egg production of older females that have been kept on their respective diets for 4 weeks (Fig. 2B). The ANOVA indicated there was no statistically significant interaction of microbiota and dietary thiamin on egg laying. These results suggest that egg production is not a function of thiamin concentration and microbiota in older females. The data further intimate that thiamin has no effect on the age-related degeneration of egg production.

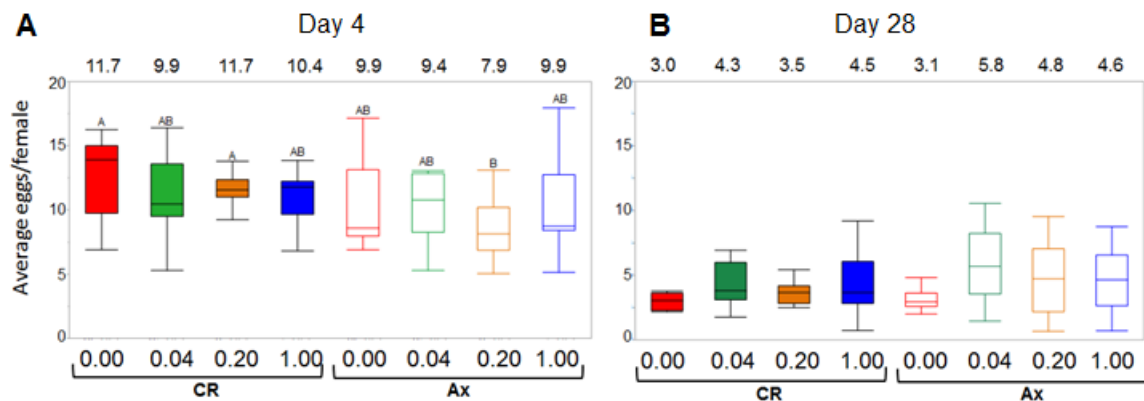


FIG 2 Impact of dietary thiamin and microbiota on egg production in adult flies. The average number of eggs/female/tube was determined at (A) 4 days and (B) 28 days on the diets, for both CR and Ax flies. Depicted are boxplots illustrating egg production. The values at the top of the graphs are the average number of eggs/female for each treatment. Treatments designated with the same letter are not significantly different. After 4 days, axenic flies raised on 0.2 μg/mL thiamin produced significantly fewer eggs (p -value < 0.0017) compared to CR flies on the same diet (p =

0.009) and the CR flies on the no-thiamin diet ($p = 0.0013$). There were no significant differences observed in egg production for any treatments after 28 days.

Thiamin is essential for offspring development and survival

Our previous assays suggested that neither the microbiota nor dietary thiamin influence egg production. However, the number of eggs produced is only an indirect measure of fecundity, as it does not quantify offspring quality, or its ability to develop from egg to adult progeny. It is therefore possible that F1 progeny of thiamin deprived flies are themselves more susceptible to lack of thiamin in the diet. We therefore hypothesized that thiamin could have an important role in *Drosophila* development, and that Ax larvae raised in a low thiamin environment would have decreased fitness, with the potential for the defect to be rescued by the transformation of the diet through the supply of thiamin by the microbiota. Hence, we examined how the microbiota and thiamin levels influence the development of eggs to adulthood on our four diets. In contrast to the egg-laying assays, this assay uncovered a substantial diet-by-microbiota interaction. Axenic eggs laid on the no-thiamin diet did not produce any pupae or adults, rather the larvae melanized and died prior to pupation (Fig. 3). However, when the microbiota was present, eggs laid on this diet developed into pupae and adults. This pattern of microbiota rescue was observed for both eggs laid by flies that were on the diets for 4 (Fig. 3A-B) and 28 (Fig. 3C-D) days, suggesting that the microbiota transforms the diet through thiamin provision to the host. These results confirm that dietary thiamin is required for larval development in Ax flies (46), and further demonstrate that microbiota derived thiamin is sufficient for development.

Nutrients not only influence the developmental success of larvae, but also their developmental rate (13). Consequently we next hypothesized that the development rate of the eggs to pupation and adulthood could be a function of microbiota and dietary thiamin. When dietary

thiamin was present, CR flies typically developed faster than their Ax counterparts on the same diet, for eggs laid either at 4 days (Fig. S2B-C, 3B) or 4 weeks of age (Fig. S2E-F, 3D). This suggests that the microbiota enhances development rate when dietary thiamin is present. For the eggs laid at 4 days, development was slower on the no-thiamin diet than any other condition, and the time to 50% pupation and adult emergence was significantly longer than all other conditions except for the Ax flies raised on the 0.04 $\mu\text{g/mL}$ thiamin medium (Fig. S2B-C, S3A, 3B). We did not observe the same developmental trend with the eggs laid at 4 weeks, as all CR flies developed at similar rates (Fig. S2E-F, S3B, 3D). Strikingly, larvae from eggs laid at 4 weeks developed faster than larvae from eggs laid at 4 days. The differences in developmental rate of the CR eggs laid at the two time-points implies that the microbiota plays a larger role in mitigating the dietary thiamin defect, as the development of eggs laid by older mothers is less constrained than the eggs laid by younger mothers. This may be a function of the greater bacterial load in older flies (3, 8, 20, 22). In contrast, Ax fly development is more contingent upon dietary thiamin, as development speed ranks with the concentration of thiamin in the medium for eggs laid at both times (Fig. S2B, E, 3B, D).

As with development speed, the interaction of dietary thiamin and microbiota promotes the survivorship of eggs to pupation and adulthood. When comparing Ax and CR eggs laid at 4 days on each diet, we observed significantly more pupation and adult emergence of CR than Ax flies on almost every diet (Fig. S2A, 3A) demonstrating the interaction of microbiota with diet elevates survival. The enhancement was a function of thiamin concentration as more thiamin typically equated to higher survivorship in the day 4 eggs (Fig 3A). For the Ax eggs laid at 4 days, thiamin was the major driver of survivorship. When observing development of the eggs laid at 28 days, there was a universal decrease in the total survivorship to pupation (Fig. S2A, S2D) and adulthood

(Fig. 3A, 3C) for all diets except for the CR flies on the 0.00 and 0.04 $\mu\text{g/mL}$ thiamin diets. On both diets, total survivorship to pupation and adulthood was increased and in disparity to the 4 day eggs, statistically equivalent to the 1 $\mu\text{g/mL}$ diet. The lack of a statistical difference in survivorship between the diets supports the idea that the presence of the microbiota has a greater influence on the eggs laid by older females as this is not dependent on thiamin concentration, rather presence of the microbiota (Fig. 3C). However, unlike with development rate, survival of Ax eggs was not proportional to thiamin concentration as the 0.2 $\mu\text{g/mL}$ thiamin diet allowed for significantly more Ax adults to develop than the 1 $\mu\text{g/mL}$ diet (Fig. 3C), suggesting there is an optimal dietary thiamin concentration to support development of Ax flies.

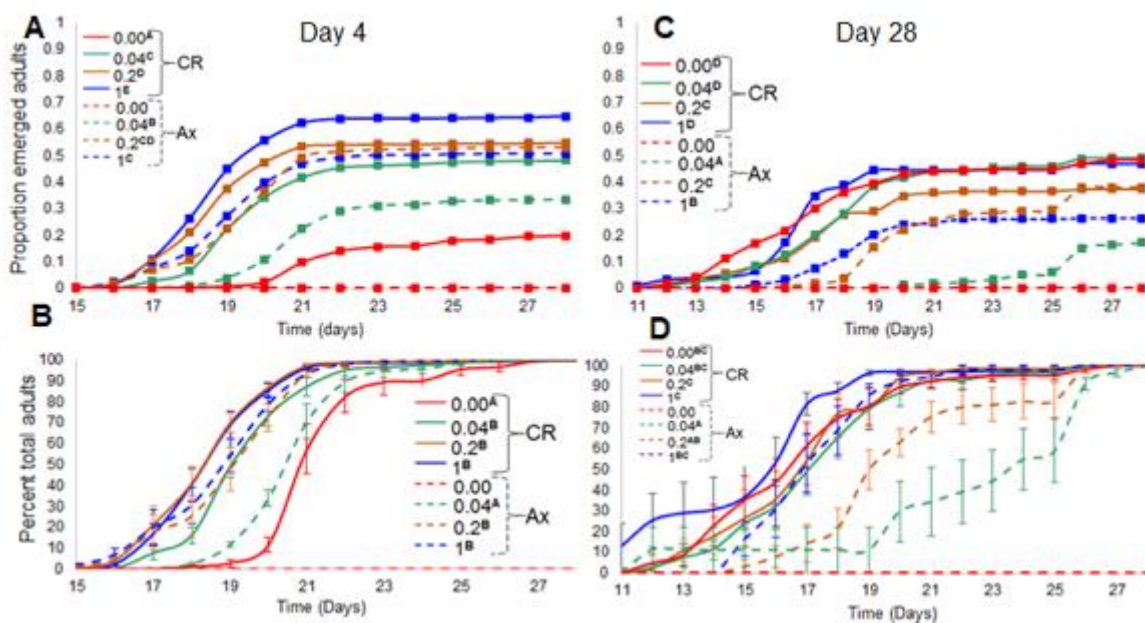


FIG 3 Impact of dietary thiamin and microbiota on fly development. Development to adulthood was assayed for CR and Ax eggs raised on diets varying in thiamin content. For all plots, treatments designated with the same letter are not significantly different from one another. (A) and (C) are survival curves depicting the proportion of emerged adults from eggs laid at 4 and 28 days, respectively. Overall, CR flies survived better than their Ax counterparts on any thiamin diet, with comparisons shown in Table S2 (4 day eggs) and Table S3 (28 day eggs). (B) and (D) show the time course for adult emergence from eggs laid by females at age 4 and 28 days, respectively. Each time point shows the \pm SEM of the proportion of the total adults emerged in a given treatment. Significant differences of the time to 50% adult emergence are displayed by the superscript lettering scheme on the legends. They correspond to the boxplots in Fig S3A for 4 day eggs and Fig S3B for 28 day eggs.

***Acetobacter pomorum* rescues development of *D. melanogaster* on a thiamin-free diet**

We next asked whether any one species in the laboratory microbiota is sufficient to rescue the developmental effects of thiamin deficiency. In our laboratory, the *D. melanogaster* microbiota is dominated by *A. pomorum*, *A. tropicalis*, *L. brevis*, and *L. plantarum*. Both *Acetobacter* spp. have the genomic potential to produce thiamin, whereas the Lactobacilli do not (15). Therefore, we hypothesized that either *Acetobacter* spp. would rescue development on a thiamin-free diet. To test this, we generated gnotobiotic flies by reassociating Ax eggs with each individual microbiota member or all 4 members on the no-thiamin diet. Only *A. pomorum* rescued development (Fig. 4). When the other bacteria were added in monoassociations, the larvae did not develop into pupae and died (Fig. S5C). Furthermore, the *A. pomorum* monoassociation was as effective at rescuing development as adding all four species in poly-association (Fig. 4). Importantly, the females that laid these eggs were reared and fed thiamin-replete media prior to oviposition, and yet their offspring were still subject to the effects of *A. pomorum*, suggesting that exogenous thiamin – whether it originates from the microbiota or diet – is more important for development than thiamin provisioned to the egg by the mother.

When dietary thiamin was present, no microbial monoassociation improved development. Development rate was not substantially affected by the presence of individual microbiota members or by having all 4 members present (Fig. S4B-C, E-F). No community member enhanced survivorship; on the contrary, Ax larvae produced significantly more pupae and adults (Fig. S4A, S4D), suggesting a deleterious effect of the microbiota in nutrient-replete conditions. For instance, *L. brevis* had a negative effect on survivorship, as gnotobiotic flies produced significantly fewer adults than their Ax counterparts (Fig. S4D). Our results therefore suggest that *L. brevis* may compete with its host for thiamin, limiting the amount available for host consumption. This

experiment confirms that only when there is a dearth of dietary thiamin, *A. pomorum* is competent to rescue development, likely due to its thiamin production capacity.

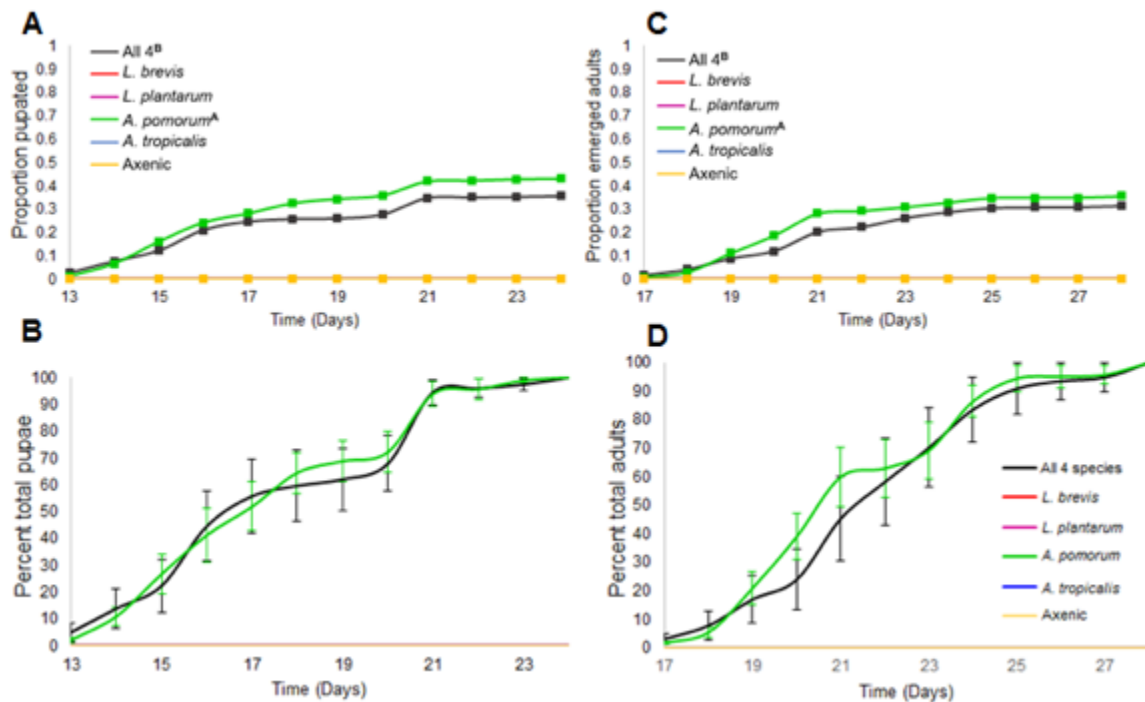


FIG 4 Impact of microbiota reassociation on fly development on the no-thiamin diet. (A) depicts the proportion of gnotobiotic eggs that have pupated and (C) depicts the proportion of eggs that emerged as adults. Significant differences from CoxPH modeling are depicted as letter superscripts next to the reassociations in the legends. The survival to pupation (A) and adulthood (C) is significantly higher in the *A. pomorum* monoassociated gnotobiotics, $p < 0.001$ in both comparisons. (B) and (D) illustrate developmental progression as the time to pupation and adult emergence, respectively, shown as \pm SEM.

***Acetobacter pomorum* associates with *D. melanogaster* on chemically defined media**

To confirm that *A. pomorum* associated with the flies in our gnotobiotic studies, adults that developed from the eggs were homogenized and plated on MRS medium, and as expected *A. pomorum* colonies were the only colonies recovered. In the poly-association, *A. pomorum* was always recovered, with *L. brevis* colonies appearing frequently, and *L. plantarum* colonies occasionally appearing as well, though not as consistently as *L. brevis*. No *A. tropicalis* colonies were recovered.

Throughout the longevity experiments (from day 5 to the end of the experiment), recently deceased CR adults on the 0.2 and 0.00 $\mu\text{g/mL}$ thiamin diets were homogenized and plated on MRS plates (Fig. S5A). In each instance, only *A. pomorum* was recovered from the flies. We also homogenized and plated the adults that developed from the eggs deposited at both 4 and 28 days. Again, the only microorganism recovered was *A. pomorum*. Our findings from the longevity experiment are consistent with the gnotobiotic experiments and show that on the no-thiamin diet, *A. pomorum* was selected for as it is able to persist, associate with, and provide thiamin to its host throughout its lifespan, and its association was unrelated to the presence of dietary thiamin.

Acetobacter pomorum* produces thiamin on the thiamin-free fly diet and enhances the growth of *Lactobacillus brevis

Our experiments suggest that *A. pomorum* was directly transforming the 0.00 $\mu\text{g/mL}$ diet through thiamin production and this was rescuing fly development. To confirm this, we measured extracellular thiamin when bacteria were cultured on the thiamin-free food. At 6 days post-inoculation, we recovered no detectable thiamin from the negative control, or from media inoculated with *A. tropicalis*, *L. brevis*, or *L. plantarum* (Fig. 5A). On average, media inoculated with *A. pomorum* yielded 24 picomoles thiamin per gram of food. When all 4 species were added in co-culture, we recovered 10-fold less thiamin, on average only 2.5 picomoles per gram of food. The difference in thiamin recovered from the *A. pomorum* monoculture compared to the community culture likely contributes to the observed lower survival of poly-association gnotobiotic flies (Fig. 4A, 4C). Additionally, or alternatively, thiamin from the microbiota may be more available to developing flies than dietary thiamin, perhaps due to lysis of bacteria in the host intestine.

The discrepancies between the mono and co-culture thiamin concentrations may be due to growth of *L. brevis* in the co-culture (Fig. 5B). Although *L. brevis* is able to grow on the food on its own, its growth was enhanced when *A. pomorum* was present, suggesting that *A. pomorum* could be providing *L. brevis* thiamin or another metabolic product (Fig. 5B). If *L. brevis* is using the thiamin produced by *A. pomorum*, this could reduce the total amount of thiamin available to the host. The number of *A. pomorum* CFUs in mono and co-culture are comparable, however, when taking into account that the starting number of cells of *A. pomorum* is 1/4 that in the monoculture, it appears that having *L. brevis* present reciprocally stimulates its growth. Both *A. tropicalis* and *L. plantarum* do not grow in the co-culture, however, *L. plantarum* produces the highest number of CFUs in monoculture (Fig. 5B). *A. tropicalis* does not grow well in monoculture, producing on average only 1,280 CFUs on the surface of the food. Taken together, these data show that the thiamin produced by *A. pomorum* is the rescuing factor, as the non-thiamin producers can grow but do not produce thiamin or rescue host development. Despite *A. pomorum* stimulating *L. brevis* growth, the fly developmental rescue is not contingent upon this.

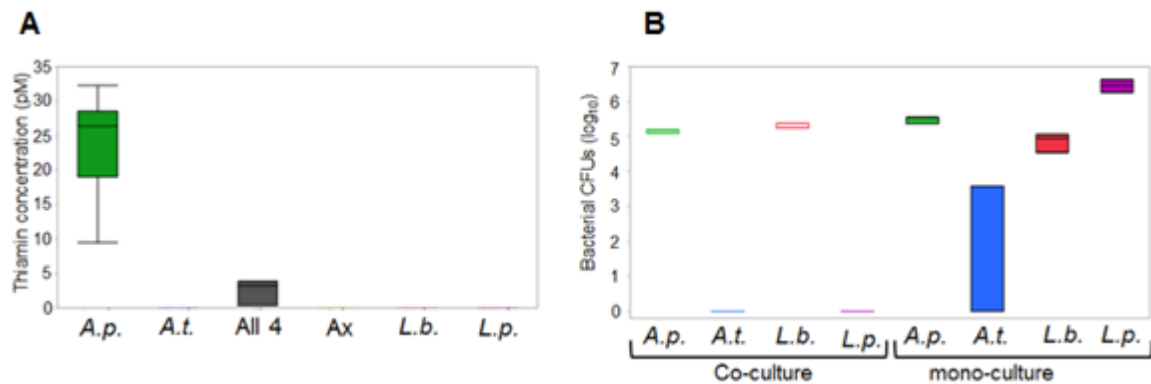


FIG 5 Production of thiamin and growth of the bacteria on the no-thiamin diet. (A) Thiamin assay results for the no-thiamin diet inoculated with *D. melanogaster* microbiota. Box plots represent the amount of thiamin produced on the no-thiamin diet by individual microbiota members, or all 4 in concert. Thiamin concentrations are shown in picomoles/gram of food. (B) CFUs recovered for each bacterium per gram of food at 3 days post inoculation. Co-culture plots represent the CFUs recovered of individual microbiota members from a mixed inoculum of all four microbiota

members. Abbreviations used for food inoculated are *A.p* for *A. pomorum*, *A.t.* for *A. tropicalis*, *L.b.* for *L. brevis*, *L.p.* for *L. plantarum* and All 4 for all microbiota members.

2.5 DISCUSSION

In the present study, we investigated the tripartite interaction between *D. melanogaster*, its microbiota, and dietary thiamin. We demonstrated that the interaction of dietary thiamin and microbiota does not have a significant impact on the life history traits of CR adult flies, as providing them a thiamin-free diet does not reduce lifespan or their ability to produce eggs. We observed a microbiota effect on the lifespan of flies, as CR flies do not live as long as Ax flies, a finding that corroborates results from previous publications (22). The major finding of this study is the impact that the interaction between diet and microbiota has on the development of flies. Ax eggs laid on a thiamin-free diet are unable to pupate and died in the larval stage, however, the presence of the microbiota in CR flies rescues their development, demonstrating that the microbiota functions to provide thiamin to its host. This developmental rescue occurs both in eggs laid by young females and eggs laid by older females, indicating the microbiota impacts the host-thiamin interaction throughout the host's lifespan. By re-associating Ax eggs with one or multiple members of their indigenous microbiota, we were able to demonstrate that *A. pomorum* is the only microbiota member capable of rescuing development despite the genomic potential for other microbiota members to synthesize thiamin. Further, we found that it is the only member capable of producing thiamin on the food.

Our finding that the dietary thiamin by microbiota interaction does not have a significant impact on *Drosophila* longevity likely has to do with how the adults were reared. The CR and Ax fly stocks were developed on standard medium containing yeast and cornmeal and both of these components contain thiamin. The larval stores of thiamin carried into adulthood likely provided

the adults with enough thiamin to survive, irrespective of their diet as adults. Similar results were found when assessing the impact of B-vitamins on the longevity of the house fly *Musca domestica* grown axenically, where depriving thiamin did not significantly reduce lifespan when compared to those fed thiamin-replete diets (48). Other holometabolous adult insects feed on low-vitamin diets, or do not eat at all, suggesting B-vitamins are superfluous for them, and our results align with this. In these insects, the critical period for acquiring B-vitamins is the larval stage, as stores can support them in adulthood, which our study indicates. The lack of a thiamin-dependent impact on Ax female egg production is in direct contrast with what has been previously reported for *D. melanogaster*. Sang and King (1961) determined that the omission of dietary thiamin reduces egg production by nearly 25% after being on the diet for 8 days, and 16 days on the diet completely eradicates production of viable eggs. Although we used different flies in our study, the major factor that likely contributes to our observed differences is the way in which the adults were raised. Sang and King raised their fly stocks on a semi-synthetic diet already low in thiamin, therefore larval stores in the egg-laying adults were likely low. The thiamin stores in flies used in our study were likely replete enough that after 4 weeks, egg output was not diminished.

The inability of Ax larvae to pupate on the no-thiamin diet demonstrates the critical nature of vitamin B₁ for development. This finding parallels with thiamin deficiencies that occur in natural populations of other animals such as alligators (49) and Great Lakes Salmonids (50), that lead to death of progeny in the early stages of their development, suggesting our system can serve as a potential model to investigate the factors that impact thiamin deficiency in other animals, including vertebrates. Unique to our study, we were able to demonstrate that the microbiota mitigates the developmental mortality caused by thiamin deficiency by supplying *Drosophila* with thiamin. The rescue of the eggs deposited by younger females on the no-thiamin diet did not attain

the same survivorship as when dietary B₁ was provided, but still allowed for pupation and emergence of adults. When dietary thiamin was provided, the microbiota enhanced both development speed and survivorship (Fig. 3A, B). The microbiota exerted a greater influence on the eggs laid by older female as CR flies on the no-thiamin diet did not have significantly less survival (comparing Fig. 3A to 3C). This improved survivorship on the no-thiamin diet is likely due to the increase in bacterial load that typically occurs with flies as they age (3, 11). As flies age, the community composition shifts, with older flies having communities dominated by *A. pomorum* (5). The higher titers of *A. pomorum* at 4 weeks likely had a two-fold effect: i) mothers were able to provision more thiamin to their eggs due to the increased numbers of the thiamin producing microbiota member, ii) mothers were able to seed the food with a higher abundance of *A. pomorum* to exert direct influence on eggs and larvae through thiamin production. Vectoring the microbiota to a food source to serve as a vitamin reservoir may be a strategy used by some species of *Drosophila*. In the wild, many *Drosophila* species feed on overripe or damaged fruits, which have a high yeast content, however, *D. suzukii* is an invasive pest which feeds on soft, marketable fruit (51). It is possible that *D. suzukii* relies on its microbiota, which it vectors to the fruit, for the supply of nutrients such as thiamin. If this holds true, targeting the microbiota may be an effective approach for pest management.

Previous studies have implicated the microbiota in providing B vitamins such as riboflavin (9, 13) and folate (14, 25) to *Drosophila*, however, this current study is the first to depict the provision of thiamin. Wong and colleagues (2014) may have failed to see a thiamin-microbiota response in their study due to using undefined diets containing yeast, which are able to supply excess thiamin to the host (13). Our results establish that *A. pomorum* is the symbiont responsible for thiamin provision to *D. melanogaster*. Gnotobiotic studies show that development of eggs on

a no-thiamin diet is rescued only when *A. pomorum* is re-associated with Ax eggs. This is congruent with the genomic potential of *A. pomorum*, as it has a complete pathway for *de novo* thiamin biosynthesis (15). Unlike the *Lactobacilli*, *A. tropicalis* also has a complete gene set for thiamin biosynthesis (15), but it was not able to grow to high densities on the food and produce thiamin. It is possible that on a different diet or in other circumstances, it too could provide thiamin for the host. In this experiment, *A. pomorum* was selected for by this particular diet. Further, we confirmed that *A. pomorum* produces thiamin on the no-thiamin fly diet. This signifies another example of an *Acetobacter* species manipulating the host diet leading to an impact on host health (18, 19). It is unclear if extracellular thiamin detected in the fly food was released by *A. pomorum* cell lysis while the bacteria grew on the food, or if it is due to the efflux of thiamin from the cell. Further, B₁ provision may also occur in the gut of *Drosophila* either by bacteria lysing in the gut, or thiamin production inside the gut. More experiments are needed to determine the exact mechanisms of how the host is acquiring thiamin from *A. pomorum*.

Acetobacter pomorum has already been recognized as an important member of the microbiota, as it regulates host signaling, affecting fly development and body size (6). It is important in maintaining host nutritional indices as flies associated with *A. pomorum* have lower TAG levels than Ax flies (16). Despite being the first symbiont shown to produce thiamin for *Drosophila*, it is not the first insect symbiont shown to produce thiamin for its host. *Rhodococcus rhodnii*, the extracellular symbiont of the Chagas disease vector *Rhodnius prolixus* produces thiamin in sterile horse blood allowing for the growth and development of its host (52). *Wigglesworthia glossinidia* and *Sodalis glossinidius* are two enteric symbionts of the tsetse fly, and in the fly, *Wigglesworthia* produces thiamin that is used by *Sodalis*. Promotion of the growth of *Sodalis* maintains host gut homeostasis and health (53). This promotion of host health through

interactions with other community members does not appear to be occurring in our system as the poly-associated gnotobiotics have reduced survival in comparison to the *A. pomorum* monoassociated flies. Although *L. brevis* growth is enhanced by *A. pomorum* and both are recovered from gnotobiotics when all 4 species are added back, *L. brevis* is able to grow on its own (Fig. 5B). The presence of *L. brevis* may be deleterious; less thiamin is being produced when it is present, and survival is reduced in monoassociated flies when dietary thiamin is present. *A. pomorum* abundance is similar in the mono and co-cultures (Fig. 5B), suggesting that the presence of *L. brevis* is accounting for the differences in thiamin production. This implies there is competition between host and auxotrophic microbiota members for thiamin, resulting in the reduction of survival in the poly-association gnotobiotics.

The evolution of symbiosis is not well-understood; what is the reckoning of costs and benefits of microbial association? Our study gives important mechanistic insight towards answering this question. Specifically, it contributes to the growing body of evidence that in *Drosophila*, microbes play a major role in vitamin provision. Our data indicate that, whilst the microbiota is deleterious for long-term health, measured by adult longevity, they are essential for offspring development in low-thiamin dietary contexts. Thus, the microbiota buffers fly fitness against dietary deficiency but limits lifespan. Altogether these data indicate that the interaction of dietary thiamin and microbiota is a mechanism of a lifespan-reproduction tradeoff.

In conclusion, we show that *D. melanogaster* microbiota, in particular *A. pomorum*, is an essential thiamin supplier in low-thiamin conditions. It transforms nutrition through thiamin production, rescuing development of axenic flies in the absence of dietary thiamin. This demonstrates the importance of the microbiota as nutritional symbionts and shows how certain environmental conditions may select for microbiota to produce specific nutrients for the host. Our

study provides a robust system to further interrogate the relationship between *D. melanogaster*, its microbiota, and other dietary components.

2.6 MATERIALS AND METHODS

***Drosophila* and bacterial stocks.** Throughout this study, *Wolbachia*-free Canton^S flies were used. CR laboratory stocks were cultured on a standard 5% (w/v) yeast, 4% (w/v) glucose, 6% (w/v) cornmeal diet at 25°C. Ax stocks were maintained on a sterile, 5% (w/v) yeast, 4% (w/v) sucrose, 6% (w/v) cornmeal diet, in sterile glass tubes at 25°C (3). For each passage, the Ax stocks were transferred aseptically into new tubes. After each transfer, flies were removed from the food after 2 days of egg laying and at least 5 flies/tube were put into sterile microfuge tubes containing 500 µL of sterile 1x PBS, homogenized, and the homogenate was plated on MRS agar, to ensure conditions were axenic. Plates were incubated at 29°C and if there was any growth, the corresponding tube of flies were discarded. *Acetobacter pomorum* DmCS_004, *Acetobacter tropicalis* DmCS_006, *Lactobacillus brevis* DmCS_003, and *Lactobacillus plantarum* DmCS_001 (15) were used in this study. *Acetobacter* spp. were grown at 30°C on MRS plates in aerobic conditions, while the *Lactobacillus* spp. were grown on MRS at 30°C in a 95% CO₂, 5% H₂ atmosphere in a Coy vinyl anaerobic chamber (Coy Laboratory Products, Inc.). When grown in liquid MRS, the *Acetobacter* spp. were grown with shaking at 225 rpm, while the *Lactobacillus* spp. were grown without shaking.

Generation of axenic and gnotobiotic flies. Ax flies were produced as previously described (16) with one amendment, we used 10% bleach (v/v) for egg dechoriation. The same protocol was followed to produce gnotobiotic flies with some modifications. Overnight cultures were pelleted and resuspended in a thiamin-free, chemically defined medium (DM4) to a final density of 10⁸

cells/mL, using the empirically derived constants for each species (16). DM4 is based on M9 medium (54), however, it is buffered by 0.1 M MES, pH 6.0. It contains 10 mM FeSO₄, 9.5 mM NH₄Cl, 0.276 mM K₂SO₄, 0.5 μM CaCl₂, 0.525 mM MgCl₂, 50 mM NaCl, 1.32 mM K₂HPO₄, 1% (v/v) Vitamin Supplement (ATCC MDVS), 1% (v/v) Trace Mineral Solution (ATCC MD-TMS), 0.1% (w/v) glucose, and 12.5% (v/v) amino acid mix based on published concentrations (55). Cell pellets were washed 2X with DM4 to remove thiamin carryover. For generating monoassociated flies, 5x10⁶ cells/vial were added to the food surface containing the sterile eggs. When all 4 species were re-associated, 1.25x10⁶ cells/vial of each species was added. Sterile DM4 was added to eggs to serve as the Ax negative control. To generate the eggs for the gnotobiotic experiments, 20 Ax females and 5 Ax males were transferred aseptically from standard medium to each tube of chemically defined diet and laid eggs for ≤18 hrs.

Chemically defined fly diet. For all experiments, we used a previously described defined diet (25), with a few minor alterations. Thiamin adsorbs to glassware (56) so precautions were taken to prevent thiamin contamination. When autoclaving was unnecessary, plasticware was used, but when glassware was necessary, it was washed with Fisherbrand Cleaning Solution (Fisher Scientific SC88-500), rinsed 10X with tap, then 10X with deionized H₂O, and baked overnight at ~200°C. Since thiamin is base-labile, a 0.1 M NaOH wash was conducted, followed by the same rinsing and baking procedure. The only amendments to the media were that 0.3 g/L cholesterol and 0.5 g/L L-cysteine HCl. Four diets were generated containing 0.00, 0.04, 0.2 or 1.0 μg/mL thiamin.

Longevity experiments. Three-day-old CR flies (20 females and 5 males) were transferred to each sterile diet tube. For each experiment, there were 3 tubes set up for each diet, and this experiment was repeated 3 times for a total of 225 flies per diet. The same design was used for Ax flies, however all transfers were conducted aseptically in a sanitized SterileGARD biosafety cabinet. All flies were passed to new sterile food every 3 to 4 days (CR flies were anesthetized with CO₂ for passage, while Ax flies were anesthetized on ice to maintain sterility). Tubes were incubated at 25°C on a 12 hr light/12 hr dark cycle and checked daily. To monitor sterility in Ax flies throughout the experiment duration, individual dead flies in each of the vials were collected aseptically and screened for microbial growth as described above. The experiments ran until every fly died, and the data for each experiment was pooled and analyzed via a Cox Proportional Hazards (CoxPH) model in SAS 9.4. The presence of microbiota, thiamin concentration, the thiamin concentration*microbiota interaction term, and replicate were fixed effects, with the tube in each experiment being a random effect. 95% Wald Confidence Limits (CL) and the point estimates were used to determine the significance of each comparison. Significant differences do not have a proportion of 1 within their CL.

Egg output experiments. Ax and CR flies were transferred to new sterile food at days 4 and 28 of the experiment where they laid eggs for ≤ 18 hrs. The eggs were counted for each tube, and the average number of eggs/female was calculated and analyzed using a fit model, which includes an Analysis of Variance (ANOVA) in JMP Pro 12.0.1. Presence of microbiota, thiamin concentration, replicate, and the B₁ concentration*microbiota cross were all fixed effects. The average eggs/female were log transformed and used as the input data. If there was a statistically significant

value from the ANOVA, pairwise comparisons were made using a *t*-test, and p-values were Bonferroni corrected as 28 comparisons were made.

Insect development. Experiments were conducted to assess the impact of dietary thiamin and microbiota on fly development. For the diet experiments, the eggs laid for the egg output experiments proceeded to develop for each diet at both egg-laying periods. There were 3 tubes for each experiment, replicated 3 times. For the gnotobiotic experiments, there were 3 tubes for 2 replicates per bacterial addition on the 0.2 µg/mL thiamin diet. For the no-thiamin diet, there were 2 replicates with 3 tubes per replicate for the addition of *L. brevis*, *L. plantarum*, and *A. tropicalis*, and 3 replicates for the Ax flies, *A. pomorum*, and all 4 species addition treatments.

Tubes were checked daily and scored for progression of the larvae. The number of pupae and subsequently adults was recorded in each tube. Replicates were pooled and developmental speed was plotted as the average of days to reach 100% pupation and 100% adult emergence. Time to 50% of all flies to reach pupation and adulthood was determined for each tube using the slope of the development rates. These values were analyzed using a fit model in JMP 12.0.1 which included an ANOVA, with treatment and replicate as the fixed effects. Confidence intervals were checked to ensure that they did not include 0. If the ANOVA was significant, pairwise comparisons were made via a *t*-test. P-values for the diet by microbiota experiments were Bonferroni corrected for the 21 comparisons. The gnotobiotic experiments on the 0.2 µg/mL thiamin diet were Bonferroni corrected for 15 comparisons.

Survival analysis of development studies. Development survival data were analyzed in R 3.4.0 using Cox Proportional Hazards models from the rms library. Time to development was modeled

as a function of experimental condition and replicate, both as unordered factors. Conditions in which no flies developed were excluded from the analyses. For scoring pupae development, eggs laid at the day-4 interval, those that did not develop into pupae by 20 days were censored, while those that did not develop into pupae by 22 days were censored for the eggs laid at 28 days. For development to pupae in the gnotobiotic experiments on the 0.2 $\mu\text{g/mL}$ thiamin diet, those that did not develop into pupae by 18 days were censored, and for the no-thiamin diet, those that did not develop by 24 days were censored. For egg development to adulthood, in all the experiments (except for the gnotobiotics on the 0.2 $\mu\text{g/mL}$ thiamin diet), those that did not develop to adults by 28 days, were censored. For the 0.2 $\mu\text{g/mL}$ thiamin diet, those that did not develop by 24 days were censored. Tukey post-hoc comparisons were conducted in R 3.4.0 using the multcomp library to identify pairwise differences between conditions. P-values were Bonferroni corrected for multiple comparisons (the development experiments of the 4 and 28 day eggs were corrected for 21 comparisons, and the B diet gnotobiotic experiments were corrected for 15 comparisons).

Microbial identification. Through the course of the longevity experiments, dead flies on the 0.2 and 0.00 $\mu\text{g/mL}$ thiamin diets were collected aseptically after the living flies were transferred to new sterile food. Individual flies were homogenized and plated on MRS agar plates as described above. Colonies were visually inspected and categorized based on morphology. Colony PCRs were conducted on multiple colonies using the 8F and 1492R universal 16S rRNA gene primers (57), and the PCR products were sequenced using the Sanger method at the Cornell University Institute of Biotechnology. Sequences were analyzed with BLAST to determine the identity of the bacteria, as well as aligned to reference 16S rRNA gene sequences from our fly microbiota isolates using Geneious 6.0.6. To determine the identity of the microbiota of the developing flies, five 1-2 day

old adults were pooled, homogenized, and plated on MRS agar as described above. Colonies were sequenced to confirm their identity. The same plating technique was used to confirm the bacteria associated with the flies in the gnotobiotic experiments. In the poly-association studies, homogenate was plated on MRS agar and incubated anoxically to select for *Lactobacillus* spp., or in the presence of oxygen and 5 µg/mL ampicillin to prevent growth of the Lactobacilli. Colony morphology was used to identify the bacteria (16).

Measurement of thiamin levels. The bacteria were added to ~7.5 mL of the no-thiamin food, in 50 mL sterile falcon tubes in the same manner as with the gnotobiotic experiments. Three tubes were used for each treatment except for *A. pomorum*, which had 6 tubes. The tubes were inspected under a dissecting microscope daily for growth. At 6 days post-inoculation the samples were processed. The concentrations of thiamin produced in the fly food media were assessed using modifications to a previously reported competitive binding assay specific for thiamin, and plates were prepared as detailed in (58). The solid fly medium was diluted in a 20% (w/v) ratio with 20/200 mM MESS (20 mM MES, pH 6.5, 200 mM sodium chloride) and incubated at ambient temperature for 5 hours with vortexing. The suspensions were centrifuged at 2,500 rpm for 5 minutes. 200 µL of supernatant was diluted with 200 µL 900 mM MES, 200 mM NaCl, pH 6.6 and the samples were thoroughly vortexed. 50 µL aliquots of supernatant were added in quadruplicate to the washed microtiter plate. Thiamin standards were prepared in the same manner using media without bacteria. Thiamin binding protein conjugated liposomes were then added to the plate and mixed as detailed in (58). The plate was processed and fluorescence measurements and data analysis were carried out as described previously (58). The critical modification to the previously reported method was the marked increase to 900 mM MES to adequately buffer the pH

change in the medium due to the acetic acid produced by *A. pomorum*. The assay performs optimally at a pH value of ~6.5, whereas the pH of the media in the presence of *A. pomorum* was 4.4 under the original 20 mM buffer conditions. Thiamin concentration was corrected for based on buffer volume and determined per gram of food.

Bacterial growth on the chemically defined diets. The bacteria were grown on the no-thiamin diet as described above for the thiamin quantification experiment, with n=3 tubes for each bacteria inoculation. Growth was inspected under a dissecting microscope daily. At 3 days, 3 mL of 1x PBS was added directly to the surface of the food. Using a Gilson P1000 pipet, the PBS was mixed by vigorously pipetting up and down. 1 mL was removed and put into a sterile microfuge tube and 1/100 and 1/1000 dilutions were made, and plated using a WASP2 Spiral Plater (Microbiology International). The monocultures and co-cultures were plated as described in the gnotobiotic section. Colonies were counted and CFUs per ml of medium determined.

FUNDING INFORMATION

For Nicolas Buchon, funding is from the Empire State Stem Cell Fund through New York State Department of Health NYSTEM contract C029542 and from NSF 1354421 and NSF 1656118, and NSF 1244378 and NSF 1354911 for Esther Angert. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

2.7 ACKNOWLEDGMENTS

We thank Cliff Kraft, Francine Arroyo, Alessandro Bonfini, and Xiao-Li Bing for stimulating discussions. We thank Franscoise Vermeulen for statistical assistance. We thank Yoni Revah and Ellery Bianco for some technical assistance.

2.8 REFERENCES

1. McFall-Ngai M, Hadfield MG, Bosch TC, Carey HV, Domazet-Lošo T, Douglas AE, Dubilier N, Eberl G, Fukami T, Gilbert SF. 2013. Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci U S A* 110:3229-3236.
2. Liu X, Hodgson JJ, Buchon N. 2017. *Drosophila* as a model for homeostatic, antibacterial, and antiviral mechanisms in the gut. *PLoS Pathog* 13:e1006277.
3. Broderick NA, Buchon N, Lemaitre B. 2014. Microbiota-induced changes in *Drosophila melanogaster* host gene expression and gut morphology. *MBio* 5:e01117-14.
4. Wong AC, Chaston JM, Douglas AE. 2013. The inconstant gut microbiota of *Drosophila* species revealed by 16S rRNA gene analysis. *ISME J* 7:1922-1932.
5. Wong CNA, Ng P, Douglas AE. 2011. Low-diversity bacterial community in the gut of the fruitfly *Drosophila melanogaster*. *Environ Microbiol* 13:1889-1900.
6. Shin SC, Kim S-H, You H, Kim B, Kim AC, Lee K-A, Yoon J-H, Ryu J-H, Lee W-J. 2011. *Drosophila* microbiome modulates host developmental and metabolic homeostasis via insulin signaling. *Science* 334:670-674.
7. Storelli G, Defaye A, Erkosar B, Hols P, Royet J, Leulier F. 2011. *Lactobacillus plantarum* promotes *Drosophila* systemic growth by modulating hormonal signals through TOR-dependent nutrient sensing. *Cell Metab* 14:403-414.
8. Ren C, Webster P, Finkel SE, Tower J. 2007. Increased internal and external bacterial load during *Drosophila* aging without life-span trade-off. *Cell Metab* 6:144-152.
9. Fridmann-Sirkis Y, Stern S, Elgart M, Galili M, Zeisel A, Shental N, Soen Y. 2014. Delayed development induced by toxicity to the host can be inherited by a bacterial-dependent, transgenerational effect. *Front Genet* 5:27.
10. Ryu J-H, Kim S-H, Lee H-Y, Bai JY, Nam Y-D, Bae J-W, Lee DG, Shin SC, Ha E-M, Lee W-J. 2008. Innate immune homeostasis by the homeobox gene caudal and commensal-gut mutualism in *Drosophila*. *Science* 319:777-782.
11. Buchon N, Broderick NA, Chakrabarti S, Lemaitre B. 2009. Invasive and indigenous microbiota impact intestinal stem cell activity through multiple pathways in *Drosophila*. *Genes Dev* 23:2333-2344.
12. Sharon G, Segal D, Ringo JM, Hefetz A, Zilber-Rosenberg I, Rosenberg E. 2010. Commensal bacteria play a role in mating preference of *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* 107:20051-20056.
13. Wong AC-N, Dobson AJ, Douglas AE. 2014. Gut microbiota dictates the metabolic response of *Drosophila* to diet. *J Exp Biol* 217:1894-1901.
14. Blatch S, Meyer KW, Harrison JF. 2010. Effects of dietary folic acid level and symbiotic folate production on fitness and development in the fruit fly *Drosophila melanogaster*. *Fly (Austin)* 4:312-319.
15. Newell PD, Chaston JM, Wang Y, Winans NJ, Sannino DR, Wong AC, Dobson AJ, Kagle J, Douglas AE. 2014. In vivo function and comparative genomic analyses of the *Drosophila* gut microbiota identify candidate symbiosis factors. *Front Microbiol* 5:576.
16. Newell PD, Douglas AE. 2014. Interspecies interactions determine the impact of the gut microbiota on nutrient allocation in *Drosophila melanogaster*. *Appl Environ Microbiol* 80:788-796.

17. Chaston JM, Dobson AJ, Newell PD, Douglas AE. 2016. Host genetic control of the microbiota mediates the *Drosophila* nutritional phenotype. *Appl Environ Microbiol* 82:671-679.
18. Chaston JM, Newell PD, Douglas AE. 2014. Metagenome-wide association of microbial determinants of host phenotype in *Drosophila melanogaster*. *MBio* 5:e01631-14.
19. Huang J-H, Douglas AE. 2015. Consumption of dietary sugar by gut bacteria determines *Drosophila* lipid content. *Biol Lett* 11:20150469.
20. Guo L, Karpac J, Tran SL, Jasper H. 2014. PGRP-SC2 promotes gut immune homeostasis to limit commensal dysbiosis and extend lifespan. *Cell* 156:109-122.
21. Buchon N, Broderick NA, Poidevin M, Pradervand S, Lemaitre B. 2009. *Drosophila* intestinal response to bacterial infection: activation of host defense and stem cell proliferation. *Cell Host Microbe* 5:200-211.
22. Clark RI, Salazar A, Yamada R, Fitz-Gibbon S, Morselli M, Alcaraz J, Rana A, Rera M, Pellegrini M, William WJ. 2015. Distinct shifts in microbiota composition during *Drosophila* aging impair intestinal function and drive mortality. *Cell Rep* 12:1656-1667.
23. Biteau B, Hochmuth CE, Jasper H. 2008. JNK activity in somatic stem cells causes loss of tissue homeostasis in the aging *Drosophila* gut. *Cell Stem Cell* 3:442-455.
24. LeBlanc JG, Milani C, de Giori GS, Sesma F, Van Sinderen D, Ventura M. 2013. Bacteria as vitamin suppliers to their host: a gut microbiota perspective. *Curr Opin Biotechnol* 24:160-168.
25. Piper MD, Blanc E, Leitão-Gonçalves R, Yang M, He X, Linford NJ, Hoddinott MP, Hopfen C, Soultoukis GA, Niemeyer C. 2014. A holidic medium for *Drosophila melanogaster*. *Nat Methods* 11:100-105.
26. Nabokina SM, Inoue K, Subramanian VS, Valle JE, Yuasa H, Said HM. 2014. Molecular identification and functional characterization of the human colonic thiamine pyrophosphate transporter. *J Biol Chem* 289:4405-4416.
27. Carini P, Campbell EO, Morré J, Sanudo-Wilhelmy SA, Thrash JC, Bennett SE, Temperton B, Begley T, Giovannoni SJ. 2014. Discovery of a SAR11 growth requirement for thiamin's pyrimidine precursor and its distribution in the Sargasso Sea. *ISME J* 8:1727-1738.
28. Sechi G, Sechi E, Fois C, Kumar N. 2016. Advances in clinical determinants and neurological manifestations of B vitamin deficiency in adults. *Nutr Rev*:nuv107.
29. Balk L, Hägerroth P-Å, Åkerman G, Hanson M, Tjärnlund U, Hansson T, Hallgrímsson GT, Zebühr Y, Broman D, Mörner T. 2009. Wild birds of declining European species are dying from a thiamine deficiency syndrome. *Proc Natl Acad Sci U S A* 106:12001-12006.
30. Balk L, Hägerroth P-Å, Gustavsson H, Sigg L, Åkerman G, Muñoz YR, Honeyfield DC, Tjärnlund U, Oliveira K, Ström K. 2016. Widespread episodic thiamine deficiency in Northern Hemisphere wildlife. *Sci Rep* 6.
31. Edwin E, Jackman R. 1970. Thiaminase I in the development of cerebrocortical necrosis in sheep and cattle. *Nature* 228:772-774.
32. Jurgenson CT, Begley TP, Ealick SE. 2009. The structural and biochemical foundations of thiamin biosynthesis. *Annu Rev Biochem* 78:569-603.
33. Breves G, Hoeller H, Harmeyer J, Martens H. 1979. Gastro-intestinal passage and balance of thiamin in sheep. *Ann Rech Vet* 10:465-466.
34. Breves G, Hoeller H, Harmeyer J, Martens H. 1980. Thiamin balance in the gastrointestinal tract of sheep. *J Anim Sci* 51:1177-1181.

35. Breves G, Brandt M, Hoeller H, Rohr K. 1981. Flow of thiamin to the duodenum in dairy cows fed different rations. *J Agric Sci* 96:587-591.
36. Said HM. 2011. Intestinal absorption of water-soluble vitamins in health and disease. *Biochem J* 437:357-372.
37. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto J-M. 2011. Enterotypes of the human gut microbiome. *Nature* 473:174-180.
38. Magnúsdóttir S, Ravcheev D, de Crécy-Lagard V, Thiele I. 2015. Systematic genome assessment of B-vitamin biosynthesis suggests co-operation among gut microbes. *Front Genet* 6:148.
39. Bakula M. 1969. The persistence of a microbial flora during postembryogenesis of *Drosophila melanogaster*. *J Invertebr Pathol* 14:365-374.
40. Leitão-Gonçalves R C-SZ, Francisco AP, Fioreze GT, Anjos M, Baltazar C, et al. 2017. Commensal bacteria and essential amino acids control food choice behavior and reproduction. *PLoS Biol* 15:e2000862.
41. Sang JH, King RC. 1961. Nutritional requirements of axenically cultured *Drosophila melanogaster* adults. *J Exp Biol* 38:793-809.
42. Lukienko P, Mel'nichenko N, Zverinskii I, Zabrodskaia S. 2000. Antioxidant properties of thiamine. *Bull Exp Biol Med* 130:874-876.
43. Iwata H. 1982. Possible role of thiamine in the nervous system. *Trends Pharmacol Sci* 3:171-173.
44. Schwenke RA, Lazzaro BP, Wolfner MF. 2016. Reproduction–immunity trade-offs in insects. *Annu Rev Entomol* 61:239-256.
45. Clifton ME, Noriega FG. 2011. Nutrient limitation results in juvenile hormone-mediated resorption of previtellogenic ovarian follicles in mosquitoes. *J Insect Physiol* 57:1274-1281.
46. Sang JH. 1956. The quantitative nutritional requirements of *Drosophila melanogaster*. *J Exp Biol* 33:45-72.
47. Robertson FW, Sang JH. 1944. The ecological determinants of population growth in a *Drosophila* culture. I. Fecundity of adult flies. *Proc R Soc Lond B Biol Sci* 132:258-277.
48. Greenberg B, Burkman AM. 1963. Effect of B-vitamins and a mixed flora on the longevity of germ-free adult houseflies, *Musca domestica* L. *J Cell Comp Physiol* 62:17-22.
49. Sepúlveda MS, Wiebe JJ, Honeyfield DC, Rauschenberger HR, Hinterkopf JP, Johnson WE, Gross TS. 2004. Organochlorine pesticides and thiamine in eggs of largemouth bass and American alligators and their relationship with early life-stage mortality. *J Wildl Dis* 40:782-786.
50. Honeyfield DC, Hinterkopf JP, Fitzsimons JD, Tillitt DE, Zajicek JL, Brown SB. 2005. Development of thiamine deficiencies and early mortality syndrome in lake trout by feeding experimental and feral fish diets containing thiaminase. *J Aquat Anim Health* 17:4-12.
51. Hamby KA, Hernández A, Boundy-Mills K, Zalom FG. 2012. Associations of yeasts with spotted-wing *Drosophila* (*Drosophila suzukii*; Diptera: Drosophilidae) in cherries and raspberries. *Appl Environ Microbiol* 78:4869-4873.
52. Harington J. 1960. Synthesis of thiamine and folic acid by *Nocardia rhodnii*, the micro-symbiont of *Rhodnius prolixus*. *Nature* 188:1027-1028.

53. Snyder AK, Deberry JW, Runyen-Janecky L, Rio RV. 2010. Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts. *Proc R Soc Lond B Biol Sci* 277:2389-2397.
54. Neidhardt FC, Bloch PL, Smith DF. 1974. Culture medium for enterobacteria. *J Bacteriol* 119:736-747.
55. Harwood CR, Cutting SM. 1990. *Molecular biological methods for Bacillus*. Wiley.
56. Farrer K, Hollenberg W. 1953. Adsorption of thiamine on glassware. *Analyst* 78:730-731.
57. Turner S, Pryer KM, Miao VP, Palmer JD. 1999. Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *J Eukaryot Microbiol* 46:327-338.
58. Edwards KA, Seog WJ, Han L, Feder S, Kraft CE, Baeumner AJ. 2016. High-Throughput Detection of Thiamine Using Periplasmic Binding Protein-Based Biorecognition. *Anal Chem* 88:8248-8256.

2.9 SUPPLEMENTARY MATERIAL

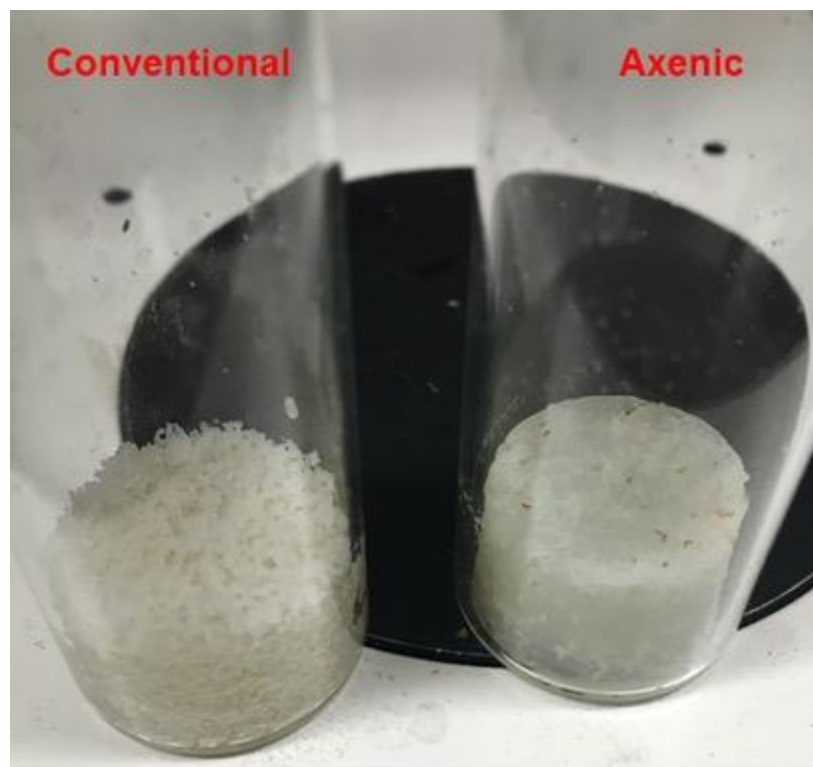


FIG S1 Larval growth of CR versus Ax flies on the no-thiamin diet. Eggs from CR flies were placed in the tube on the left and eggs from Ax flies were laid in the tube on the right. As the CR larvae grew, they churned up the food as they eat. The food in the Ax tube is not as well churned as the CR tube, the medium appears to be drying out, and dead melanized larvae are visible on the food surface.

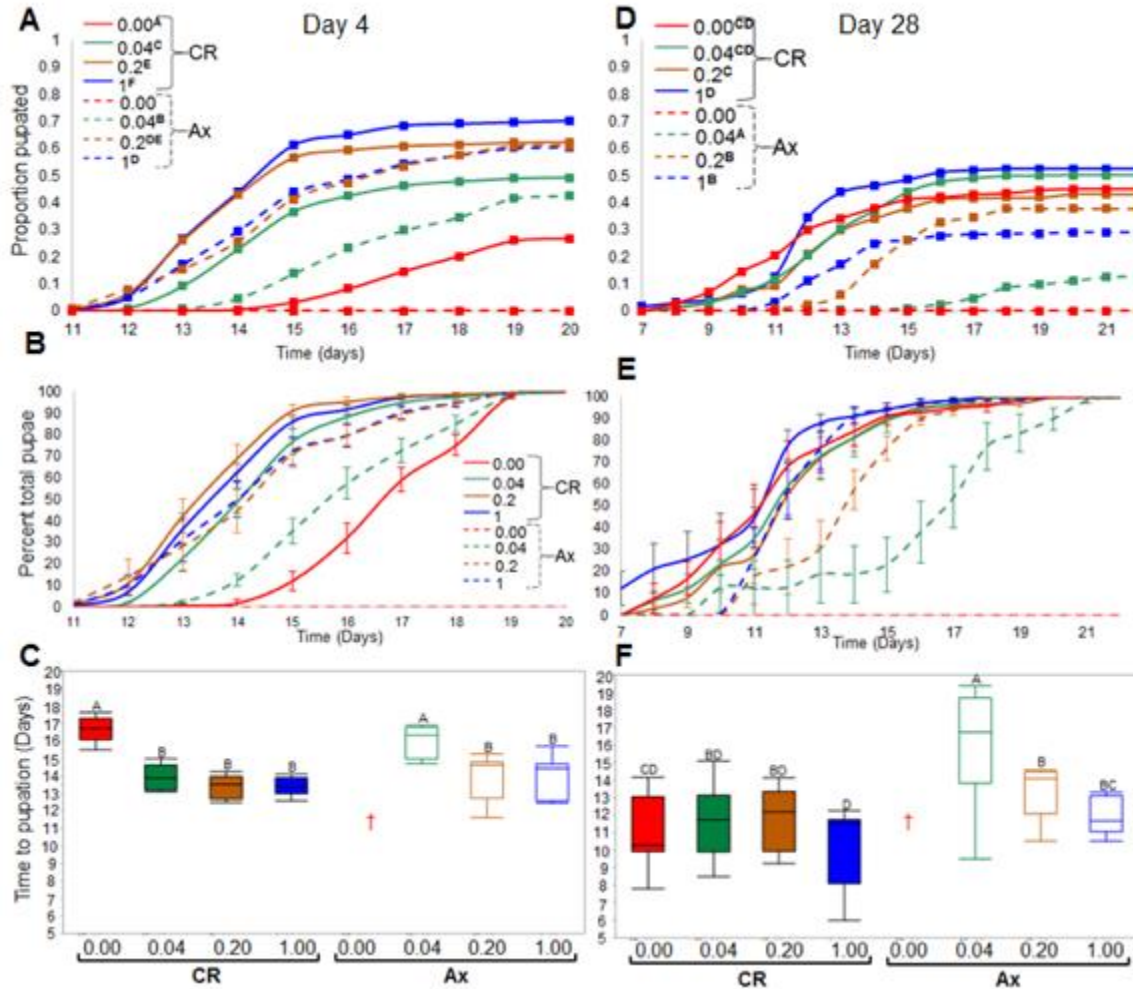
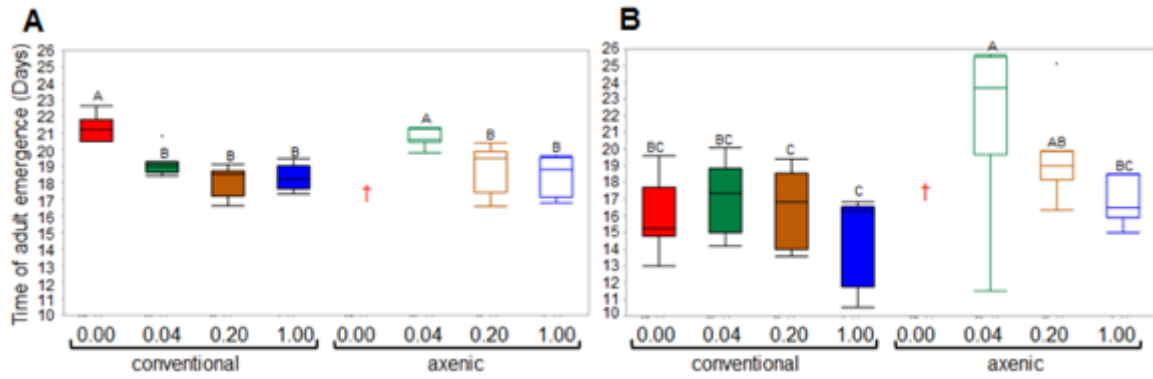


FIG S2 Impact of dietary thiamin and microbiota on fly development. Development to pupation was assayed for CR and Ax eggs raised on diets varying in thiamin content. For all plots, treatments in legends designated with the same letter are not significantly different from one another. (A) and (D) are survival curves depicting the proportion of pupae from eggs laid at 4 and 28 days, respectively. Overall, CR flies survived better than their Ax counterparts on any thiamin diet, with comparisons in Table S4 (4 day eggs) and Table S5 (28 day eggs). (B) and (E) time for 100% of progeny to reach pupation from eggs laid by females at 4 and 28 days, respectively. Each time point shows the \pm SEM of the proportion of progeny to pupate in a given treatment. (C) and (F) are box plots representing the time of emergence of 50% of the total population of pupae from the eggs laid at 4 days and 28 days, respectively (t -test, p -value < 0.0023). The \dagger represents no survival to adulthood. In (C), the 0.00 $\mu\text{g/mL}$ diet was significantly slower than all the other conventional diets and the 0.2 $\mu\text{g/mL}$ and 1 $\mu\text{g/mL}$ axenic diets with $p < 0.0001$ for each comparison. The 0.04 $\mu\text{g/mL}$ thiamin diet was significantly slower than all the other conventional diets than what was previously mentioned, and all the other axenic diets $p < 0.0001$ for all comparisons. (F) The 0.04 $\mu\text{g/mL}$ diet developed significantly slower than all the other conventional and axenic diets with $p < 0.0001$ for all comparisons. The 0.2 $\mu\text{g/mL}$ axenic diet was slower than the 1 $\mu\text{g/mL}$ and 0.00 $\mu\text{g/mL}$ conventional diets, $p < 0.0001$ and $p = 0.001$ respectively, and the 1 $\mu\text{g/mL}$ axenic developed slower than its conventional counterpart $p = 0.0021$.



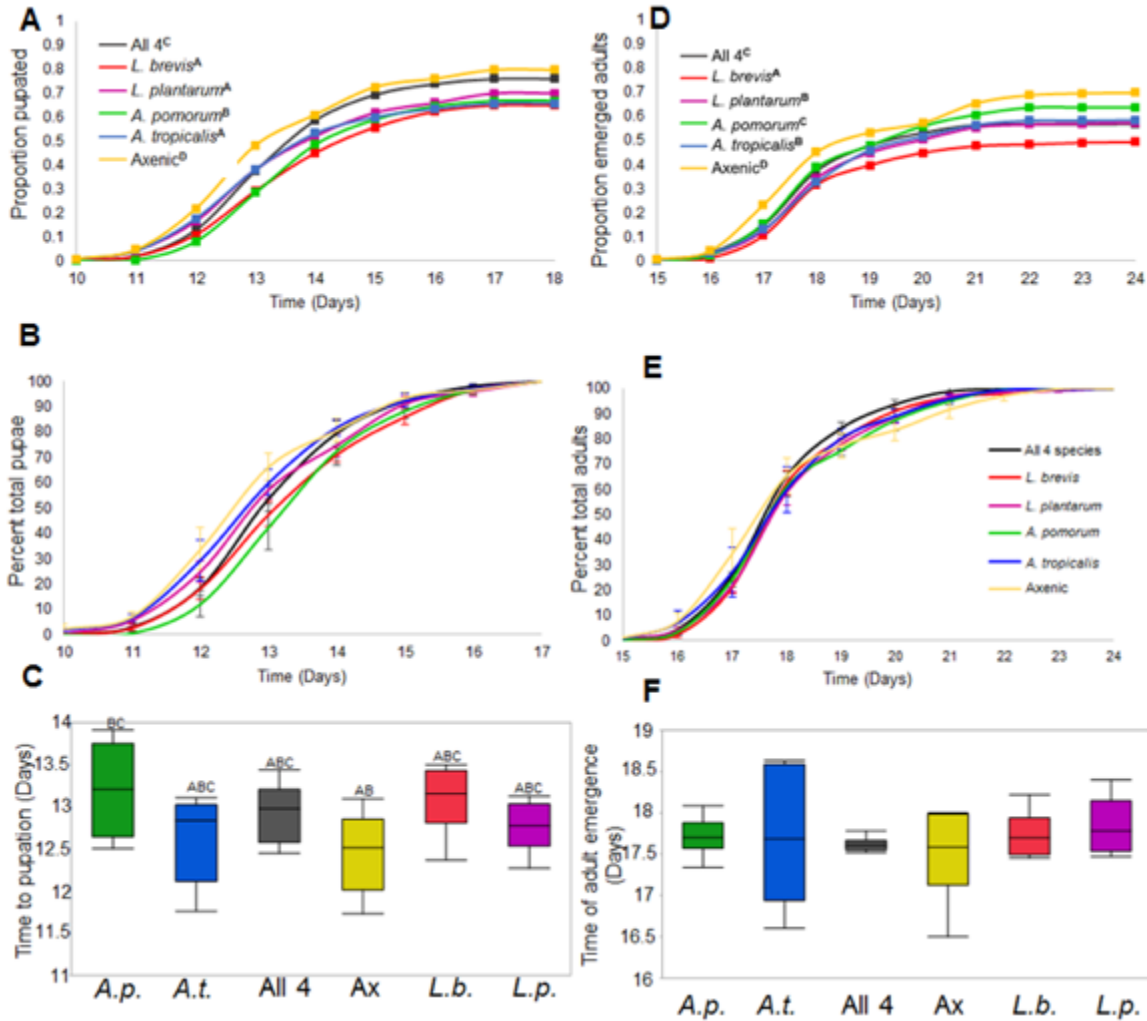


FIG S4 Impact of microbiota on fly development on the 0.2 $\mu\text{g/mL}$ thiamin diet. (A) depicts the proportion of gnotobiotic eggs that have pupated and (D) depicts the proportion of eggs that emerged as adults. Significant differences from CoxPH modelling are displayed as letter superscripts next to the reassociations in the legends. (A) for all statistically significant pairwise comparisons of proportion pupated, $p < 0.001$, except for *L. plantarum* versus axenic, $p = 0.00131$. (D) for all statistically significant pairwise comparisons of proportion adult emergence, $p < 0.001$, except for *L. brevis* versus *A. pomorum*, $p = 0.00206$, *A. tropicalis* $p = 0.02282$, all 4 species $p = 0.00328$, and *L. plantarum*, $p = 0.03303$, and axenic versus *A. tropicalis*, $p = 0.00168$.

(B) and (E) show the development as the average time to pupation (B) and adult emergence (E). Each time point shows the average proportion of eggs to pupate (B) or adults to emerge (E), with \pm standard error bars, $n = 6$ for each point. (C) and (F) are box plots representing the time of pupation (C) and adult emergence of 50% of the total population (F). The *A. pomorum* gnotobiotic pupae develop significantly slower than the axenic pupae $p = 0.0024$. (F) There were no statistically significant differences in speed of development between samples for both pupae formation and adult eclosion.

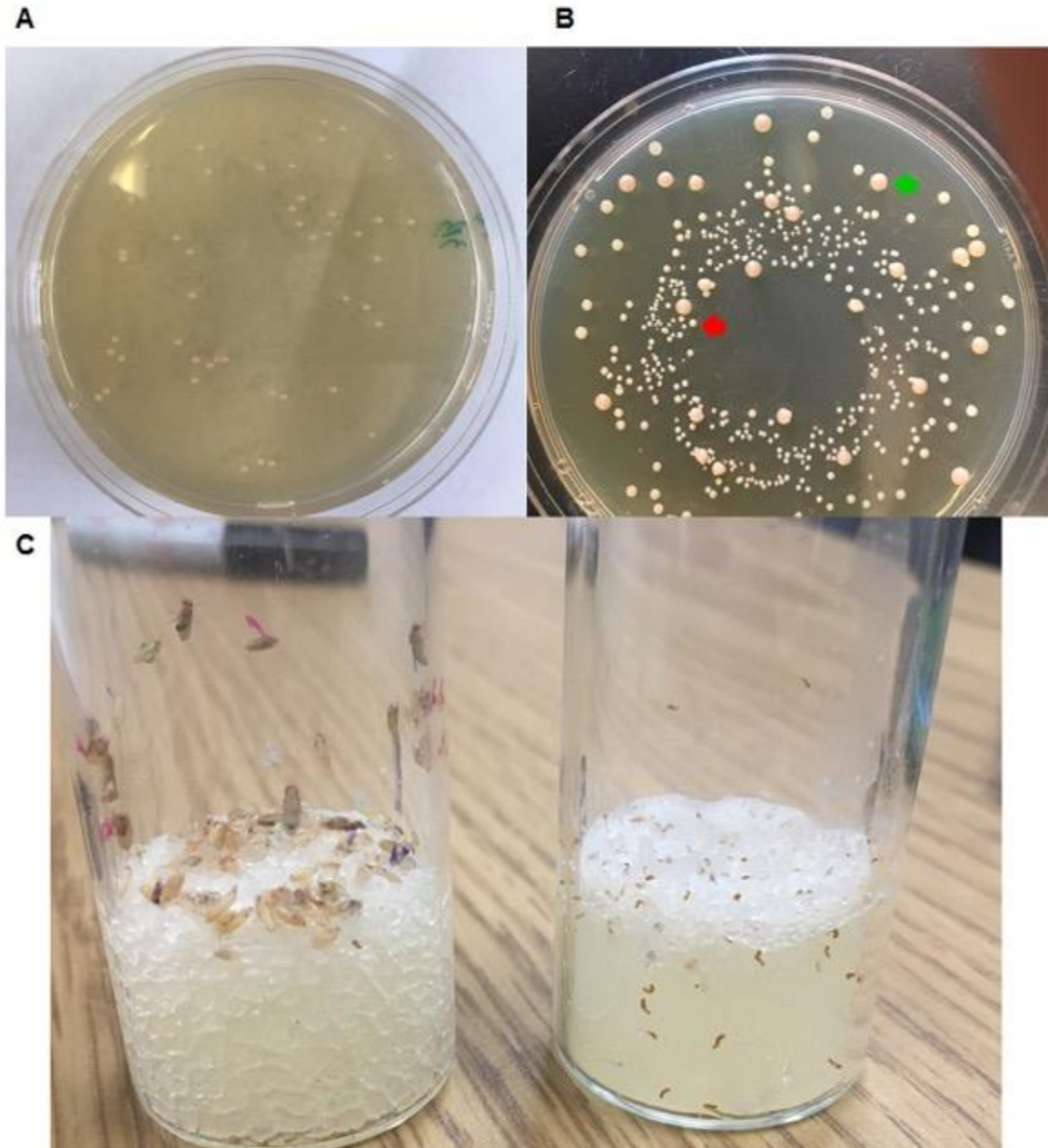


FIG S5 Recovery of bacteria from conventional and gnotobiotic flies on the no-thiamin diet, and gnotobiotic fly development on the no thiamin diet.

(A) is an MRS agar plate inoculated with 100 μ L of fly homogenate grown on the no-thiamin diet. All colonies on the plate are *A. pomorum*. (B) depicts an MRS agar plate spiral plated with 50 μ L of the homogenate of 5 adult gnotobiotic flies reassociated with the four bacterial species. There are 2 colony types present on the plate, the larger tan tinted colonies are *A. pomorum* (green arrow), while the smaller white colonies are *L. brevis* (red arrow). (C) Gnotobiotic fly development on the no-thiamin diet. The tube to the left depicts gnotobiotic flies mono-associated with *A. pomorum*. Pupae and adults are present. The tube to the right shows gnotobiotic flies associated with *L. brevis*, with only the dead melanized larvae visible.

TABLE S1 CoxPH comparisons of longevity experiment

Comparison	Point estimate	95% Wald Confidence Limits	95% Wald Confidence Limits
0 diet comparison	1.552	0.947	2.544
0.04 diet comparison	1.72	1.049	2.819
0.2 diet comparison	2.37	1.444	3.89
1 diet comparison	2.107	1.285	3.456
0 vs 0.04 CR	1.279	0.78	2.096
0 vs 0.2 CR	0.912	0.557	1.494
0 vs 1 CR	0.941	0.574	1.541
0.04 vs 0.2 CR	0.713	0.435	1.169
0.04 vs 1 CR	0.736	0.449	1.205
0.2 vs 1 CR	1.031	0.629	1.689
0 vs 0.04 Ax	1.417	0.865	2.322
0 vs 0.2 Ax	1.393	0.85	2.285
0 vs 1 Ax	1.277	0.78	2.092
0.04 vs 0.2 Ax	0.983	0.6	1.612
0.04 vs 1 Ax	0.901	0.55	1.477
0.2 vs 1 Ax	0.913	0.559	1.502

Table S1 Included are the point estimates and upper and lower confidence limits for the pairwise comparisons. Significant differences do not have a value of 1 within the confidence limits and they are highlighted in yellow.

TABLE S2 CoxPH comparisons of survival to adulthood of 4 day eggs

	Estimate	Std. Error	z value	Pr(> z)	
0.04µg/mL thiamin - 0.00µg/mL thiamin	1.16803	0.06017	19.413	< 0.001	***
0.2µg/mL thiamin - 0.00µg/mL thiamin	1.42494	0.0578	24.653	< 0.001	***
1µg/mL thiamin - 0.00µg/mL thiamin	1.73016	0.05723	30.231	< 0.001	***
ax 0.04µg/mL thiamin - 0.00µg/mL thiamin	0.58732	0.06492	9.047	< 0.001	***
ax 0.2µg/mL thiamin - 0.00µg/mL thiamin	1.29136	0.06129	21.07	< 0.001	***
ax 1µg/mL thiamin - 0.00µg/mL thiamin	1.25635	0.05956	21.095	< 0.001	***
0.2µg/mL thiamin - 0.04µg/mL thiamin	0.2569	0.0458	5.609	< 0.001	***
1µg/mL thiamin - 0.04µg/mL thiamin	0.56213	0.04491	12.516	< 0.001	***
ax 0.04µg/mL thiamin - 0.04µg/mL thiamin	-0.58071	0.0545	-10.656	< 0.001	***
ax 0.2µg/mL thiamin - 0.04µg/mL thiamin	0.12333	0.05002	2.466	0.16803	
ax 1µg/mL thiamin - 0.04µg/mL thiamin	0.08832	0.04795	1.842	0.51337	
1µg/mL thiamin - 0.2µg/mL thiamin	0.30523	0.04155	7.346	< 0.001	***
ax 0.04µg/mL thiamin - 0.2µg/mL thiamin	-0.83761	0.05176	-16.184	< 0.001	***
ax 0.2µg/mL thiamin - 0.2µg/mL thiamin	-0.13358	0.04709	-2.837	0.06647	.
ax 1µg/mL thiamin - 0.2µg/mL thiamin	-0.16859	0.04472	-3.77	0.00299	**
ax 0.04µg/mL thiamin - 1µg/mL thiamin	-1.14284	0.05115	-22.341	< 0.001	***
ax 0.2µg/mL thiamin - 1µg/mL thiamin	-0.4388	0.04635	-9.467	< 0.001	***
ax 1µg/mL thiamin - 1µg/mL thiamin	-0.47381	0.04402	-10.764	< 0.001	***
ax 0.2µg/mL thiamin - ax 0.04µg/mL thiamin	0 0.70404	0.05568	12.645	< 0.001	***
ax 1µg/mL thiamin - ax 0.04µg/mL thiamin	0.66902	0.05374	12.449	< 0.001	***
ax 1µg/mL thiamin - ax 0.2µg/mL thiamin	-0.03501	0.04925	-0.711	0.99175	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

TABLE S3 CoxPH comparisons of survival to adulthood of 28 day eggs

	Estimate	Std. Error	z value	Pr(> z)	
0.04µg/mL thiamin - 0.00µg/mL thiamin	-4.35E-05	8.64E-02	-0.001	1	
0.2µg/mL thiamin - 0.00µg/mL thiamin	-3.75E-01	9.53E-02	-3.937	0.00158	**
1µg/mL thiamin - 0.00µg/mL thiamin	-1.83E-02	8.79E-02	-0.208	0.99999	
ax 0.04µg/mL thiamin - 0.00µg/mL thiamin	-1.46E+00	1.03E-01	-14.129	< 0.001	***
ax 0.2µg/mL thiamin - 0.00µg/mL thiamin	-5.64E-01	8.80E-02	-6.407	< 0.001	***
ax 1µg/mL thiamin - 0.00µg/mL thiamin	-9.22E-01	1.05E-01	-8.765	< 0.001	***
0.2µg/mL thiamin - 0.04µg/mL thiamin	-3.75E-01	8.91E-02	-4.21	< 0.001	***
1µg/mL thiamin - 0.04µg/mL thiamin	-1.83E-02	8.12E-02	-0.225	0.99999	
ax 0.04µg/mL thiamin - 0.04µg/mL thiamin	-1.46E+00	9.88E-02	-14.773	< 0.001	***
ax 0.2µg/mL thiamin - 0.04µg/mL thiamin	-5.64E-01	8.32E-02	-6.774	< 0.001	***
ax 1µg/mL thiamin - 0.04µg/mL thiamin	-9.22E-01	1.02E-01	-9.078	< 0.001	***
1µg/mL thiamin - 0.2µg/mL thiamin	3.57E-01	9.09E-02	3.929	0.0016	**
ax 0.04µg/mL thiamin - 0.2µg/mL thiamin	-1.08E+00	1.06E-01	-10.223	< 0.001	***
ax 0.2µg/mL thiamin - 0.2µg/mL thiamin	-1.88E-01	9.13E-02	-2.061	0.37238	
ax 1µg/mL thiamin - 0.2µg/mL thiamin	-5.47E-01	1.08E-01	-5.058	< 0.001	***
ax 0.04µg/mL thiamin - 1µg/mL thiamin	-1.44E+00	9.95E-02	-14.489	< 0.001	***
ax 0.2µg/mL thiamin - 1µg/mL thiamin	-5.45E-01	8.36E-02	-6.526	< 0.001	***
ax 1µg/mL thiamin - 1µg/mL thiamin	-9.04E-01	1.02E-01	-8.896	< 0.001	***
ax 0.2µg/mL thiamin - ax 0.04µg/mL thiamin	8.96E-01	9.84E-02	9.104	< 0.001	***
ax 1µg/mL thiamin - ax 0.04µg/mL thiamin	5.37E-01	1.14E-01	4.719	< 0.001	***
ax 1µg/mL thiamin - ax 0.2µg/mL thiamin	-3.58E-01	9.96E-02	-3.6	0.00577	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

TABLE S4 CoxPH comparisons of survival to pupation of 4 day eggs

	Estimate	Std. Error	z value	Pr(> z)	
0.04µg/mL thiamin - 0.00µg/mL thiamin	0.952	0.05429	17.537	<0.001	***
0.2µg/mL thiamin - 0.00µg/mL thiamin	1.3331	0.05081	26.235	<0.001	***
1µg/mL thiamin - 0.00µg/mL thiamin	1.55603	0.05068	30.703	<0.001	***
ax 0.04µg/mL thiamin - 0.00µg/mL thiamin	0.54928	0.05639	9.741	<0.001	***
ax 0.2µg/mL thiamin - 0.00µg/mL thiamin	1.21129	0.05423	22.337	<0.001	***
ax 1µg/mL thiamin - 0.00µg/mL thiamin	1.19726	0.05214	22.964	<0.001	***
0.2µg/mL thiamin - 0.04µg/mL thiamin	0.3811	0.04421	8.62	<0.001	***
1µg/mL thiamin - 0.04µg/mL thiamin	0.60404	0.04386	13.773	<0.001	***
ax 0.04µg/mL thiamin - 0.04µg/mL thiamin	-0.40272	0.0505	-7.974	<0.001	***
ax 0.2µg/mL thiamin - 0.04µg/mL thiamin	0.25929	0.04794	5.409	<0.001	***
ax 1µg/mL thiamin - 0.04µg/mL thiamin	0.24526	0.04563	5.375	<0.001	***
1µg/mL thiamin - 0.2µg/mL thiamin	0.22293	0.0394	5.658	<0.001	***
ax 0.04µg/mL thiamin - 0.2µg/mL thiamin	-0.78382	0.04674	-16.77	<0.001	***
ax 0.2µg/mL thiamin - 0.2µg/mL thiamin	-0.12181	0.04408	-2.764	0.0819	.
ax 1µg/mL thiamin - 0.2µg/mL thiamin	-0.13584	0.0414	-3.281	0.0178	*
ax 0.04µg/mL thiamin - 1µg/mL thiamin	-1.00675	0.04656	-21.624	<0.001	***
ax 0.2µg/mL thiamin - 1µg/mL thiamin	-0.34475	0.04377	-7.877	<0.001	***
ax 1µg/mL thiamin - 1µg/mL thiamin	-0.35877	0.04114	-8.721	<0.001	***
ax 0.2µg/mL thiamin - ax 0.04µg/mL thiamin	0 0.66201	0.05043	13.126	<0.001	***
ax 1µg/mL thiamin - ax 0.04µg/mL thiamin	0.64798	0.04818	13.45	<0.001	***
ax 1µg/mL thiamin - ax 0.2µg/mL thiamin	-0.01403	0.04555	-0.308	0.9999	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

TABLE S5 CoxPH comparisons of survival to pupation of 28 day eggs

	Estimate	Std. Error	z value	Pr(> z)	
0.04µg/mL thiamin - 0.00µg/mL thiamin	0.12992	0.08787	1.479	0.75336	
0.2µg/mL thiamin - 0.00µg/mL thiamin	-0.10129	0.09353	-1.083	0.93163	
1µg/mL thiamin - 0.00µg/mL thiamin	0.23258	0.08765	2.653	0.10778	
ax 0.04µg/mL thiamin - 0.00µg/mL thiamin	-1.65024	0.1139	-14.488	< 0.001	***
ax 0.2µg/mL thiamin - 0.00µg/mL thiamin	-0.47331	0.0898	-5.271	< 0.001	***
ax 1µg/mL thiamin - 0.00µg/mL thiamin	-0.71718	0.10381	-6.909	< 0.001	***
0.2µg/mL thiamin - 0.04µg/mL thiamin	-0.23122	0.08495	-2.722	0.09043	.
1µg/mL thiamin - 0.04µg/mL thiamin	0.10266	0.07835	1.31	0.84431	
ax 0.04µg/mL thiamin - 0.04µg/mL thiamin	-1.78016	0.1081	-16.468	< 0.001	***
ax 0.2µg/mL thiamin - 0.04µg/mL thiamin	-0.60323	0.08295	-7.272	< 0.001	***
ax 1µg/mL thiamin - 0.04µg/mL thiamin	-0.84711	0.09831	-8.617	< 0.001	***
1µg/mL thiamin - 0.2µg/mL thiamin	0.33387	0.08504	3.926	0.00164	**
ax 0.04µg/mL thiamin - 0.2µg/mL thiamin	-1.54895	0.1122	-13.805	< 0.001	***
ax 0.2µg/mL thiamin - 0.2µg/mL thiamin	-0.37202	0.08778	-4.238	< 0.001	***
ax 1µg/mL thiamin - 0.2µg/mL thiamin	-0.61589	0.10216	-6.029	< 0.001	***
ax 0.04µg/mL thiamin - 1µg/mL thiamin	-1.88282	0.10743	-17.525	< 0.001	***
ax 0.2µg/mL thiamin - 1µg/mL thiamin	-0.70589	0.08158	-8.653	< 0.001	***
ax 1µg/mL thiamin - 1µg/mL thiamin	-0.94976	0.09688	-9.803	< 0.001	***
ax 0.2µg/mL thiamin - ax 0.04µg/mL thiamin	1.17693	0.10823	10.875	< 0.001	***
ax 1µg/mL thiamin - ax 0.04µg/mL thiamin	0.93306	0.11995	7.779	< 0.001	***
ax 1µg/mL thiamin - ax 0.2µg/mL thiamin	-0.24387	0.09666	-2.523	0.14742	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

CHAPTER 3

GENOMIC INSIGHTS INTO THE THIAMIN METABOLISM OF *PAENIBACILLUS THIAMINOLYTICUS* NRRL B-4156 AND *P. APIARIUS* NRRL B-23460[#]

3.1 ABSTRACT

Paenibacillus thiaminolyticus is the model organism for studying thiaminase I, an enigmatic extracellular enzyme. Originally isolated from the feces of clinical patients suffering from thiamin deficiency, *P. thiaminolyticus* has been implicated in thiamin deficiencies in humans and other animals due to its ability to produce this thiamin-degrading enzyme. Its close relative, *P. apiarius*, also produces thiaminase I and was originally isolated from dead honeybee larvae, though it has not been reported to be a honeybee pathogen. We generated draft genomes of the type strains of both species, *P. thiaminolyticus* NRRL B-4156 and *P. apiarius* NRRL B-23460, to deeply explore potential routes of thiamin metabolism. We discovered that the thiaminase I gene is located in a highly conserved operon with thiamin biosynthesis and salvage genes, as well as genes involved in the biosynthesis of the antibiotic bacimethrin. Based on metabolic pathway predictions, *P. apiarius* NRRL B-23460 has the genomic capacity to synthesize thiamin *de novo* using a pathway that is rarely seen in bacteria, but *P. thiaminolyticus* NRRL B-4156 is a thiamin auxotroph. Both genomes encode importers for thiamin and the pyrimidine moiety of thiamin, as well as enzymes to synthesize thiamin from pyrimidine and thiazole.

3.2 ABBREVIATIONS

Glyceraldehyde-3-phosphate (G3P), phospho5-aminoimidazole ribotide (AIR), thiamin monophosphate (TMP), thiamin pyrophosphate (TPP), hydroxymethyl pyrimidine (HMP), hydroxymethyl pyrimidine phosphate (HMP-P), hydroxymethyl pyrimidine pyrophosphate

(HMP-PP), hydroxymethyl pyrimidine-organic nucleophile (HMP*), thiazole carboxylate (THZ), thiazole phosphate carboxylate (THZ-P), thiamin pyrophosphokinase (TPK)

3.3 INTRODUCTION

Prior to World War II, beriberi and other vitamin deficiencies were prevalent in Japan and linked to a diet composed almost entirely of polished rice [1]. Additionally, it was discovered that certain fish and shellfish contained no thiamin and moreover any thiamin added to these raw foodstuffs was quickly destroyed [2]. While investigating potential links between the intestinal microbiota and beriberi, Shibata and colleagues found that when thiamin was added to feces or infused in the colon of patients suffering thiamin deficiency, the added thiamin disappeared [2, 3]. The thiaminase enzyme responsible for the destruction of thiamin in feces and in animal tissues was discovered shortly thereafter. Several bacteria, including *Paenibacillus thiaminolyticus*, were isolated by Matsukawa and Misawa from patient fecal samples with thiaminase activity [2]. The discovery of thiaminase producing bacteria facilitated extensive research efforts to understand the biochemistry of thiaminase and the biology of *P. thiaminolyticus* [4].

Paenibacillus thiaminolyticus became a model system for studying the secreted bacterial thiaminase now known as thiaminase I [5-10]. Thiaminase I catalyzes the base substitution of the thiazole moiety of thiamin with numerous organic nucleophiles such as pyridine, quinolone, or compounds containing a sulfhydryl group, like cysteine [2, 10, 11]. Early studies of this extracellular enzyme found that thiaminase I activity is repressed when high concentrations of thiamin are added to cultures and culture supernatant [8, 9]. The crystal structure of *P. thiaminolyticus* thiaminase I revealed that the 42 kDa protein has a catalytic cysteine residue and the protein is structurally similar to the group II periplasmic binding proteins, particularly the

thiamin-binding protein TbpA in *E. coli* [12]. We recently found that *Paenibacillus apiarius* also has thiaminase I activity (unpublished). This close relative of *P. thiaminolyticus* was originally isolated from the larvae of dead honeybees, although it was not the causative agent of their death [13]. Despite the extensive biochemical and mechanistic understanding of the enzyme, the biological function and context in which *P. apiarius*, *P. thiaminolyticus* and other thiaminase I producers use thiaminase I remains a mystery [14].

Although thiaminase I activity is found in plants such as bracken fern [15] and nardoo [16], as well as in animals such as crustaceans, ruminants, and fish, the only confirmed producers of thiaminase I are microbial, including one eukaryote, the amoeba *Naegleria gruberi* [15, 17, 18]. Thiaminase I activity in food contributes to thiamin deficiency in animals and is implicated in Early Mortality Syndrome in salmonids in the Great Lakes and Baltic Sea [18]. A link between *P. thiaminolyticus* and this thiamin deficiency syndrome has been suggested, as *P. thiaminolyticus* has been isolated from the viscera of alewife, a fish with high thiaminase activity that is a food source for Great Lakes salmonids. Additionally, it was demonstrated that Early Mortality Syndrome could be induced in lake trout fed an experimental diet supplemented with *P. thiaminolyticus* [18, 19]. As with humans, *P. thiaminolyticus* is not always isolated from intestinal contents of fish with high thiaminase I activity so other sources of the enzyme likely impact thiamin metabolism in populations of animals [20].

Thiaminase I enzymes are not widely distributed in the microbial world and are produced by a small subset of phylogenetically diverse microorganisms. By sequencing the genomes of the type strains, *P. thiaminolyticus* NRRL B-4156 and its relative *P. apiarius* NRRL B-23460, we aim to establish the genomic context of the thiaminase I gene to help gain a better understanding of the biological function of the enzyme. The draft genomes have helped uncover the routes of vitamin

B1 metabolism available to these bacteria, which will help inform our model of the ecological role of thiaminase I, and perhaps its contribution to vitamin deficiencies in animals.

3.4 ORGANISM INFORMATION

Classification and features

The original isolate of *P. thiaminolyticus*, classified as *Bacillus thiaminolyticus*, was obtained from the feces of a Japanese patient suffering from thiamin deficiency and chronic constipation [2]. Additional strains of *P. thiaminolyticus* have been isolated from fecal samples of healthy human subjects from Kyoto and Ube City, as well as those with symptoms of thiamin deficiency [2, 4]. Aside from being associated with human feces, *P. thiaminolyticus* reportedly induced bacteremia in an 80-year-old hospital patient undergoing hemodialysis for end-stage renal disease [21]. Strains of *P. thiaminolyticus* have been found in the alimentary tract and feces of thiamin deficient lambs, ewes, and sheep [22], and from the viscera of Lake Michigan alewives [18, 23]. Other isolates have been recovered from honeybees [24] and from soil [4]. Growth of *P. thiaminolyticus* on defined minimal media requires the addition of thiamin or the two moieties that form thiamin [6]. Like some strains of *P. thiaminolyticus*, *P. apiarius* was isolated from dead honeybee larvae, adults, and honeycombs [13]. It is not suspected to be a honeybee pathogen as *P. apiarius* spores fed to larvae and adults did not induce death or any obvious pathology [13]. A few *P. thiaminolyticus* strains have been erroneously classified as *P. apiarius* [25]. In contrast to *P. thiaminolyticus*, *P. apiarius* has not been studied extensively.

Both species are rod-shaped endospore formers and produce a single ellipsoid endospore in a swollen sporangium, with the spore coat of *P. apiarius* described as unusually thick [25]. The spore produced by *P. apiarius* has a rectangular outline, unlike the more ellipsoid shape seen in *P.*

thiaminolyticus [25]. *P. apiarius* cells are slightly larger than *P. thiaminolyticus* cells as they range from 3.0-5.0 μm in length and 0.7-0.8 μm in width, while *P. thiaminolyticus* cells are 2.0-3.0 μm long and 0.5-1.0 μm wide [24, 25] (Fig. 1). The predominant cellular fatty acid in both *P. apiarius* and *P. thiaminolyticus* is anteiso- $\text{C}_{15:0}$ [25], and both have a Gram-positive cell wall. General features of the two organisms are summarized in Tables 1A and 1B.

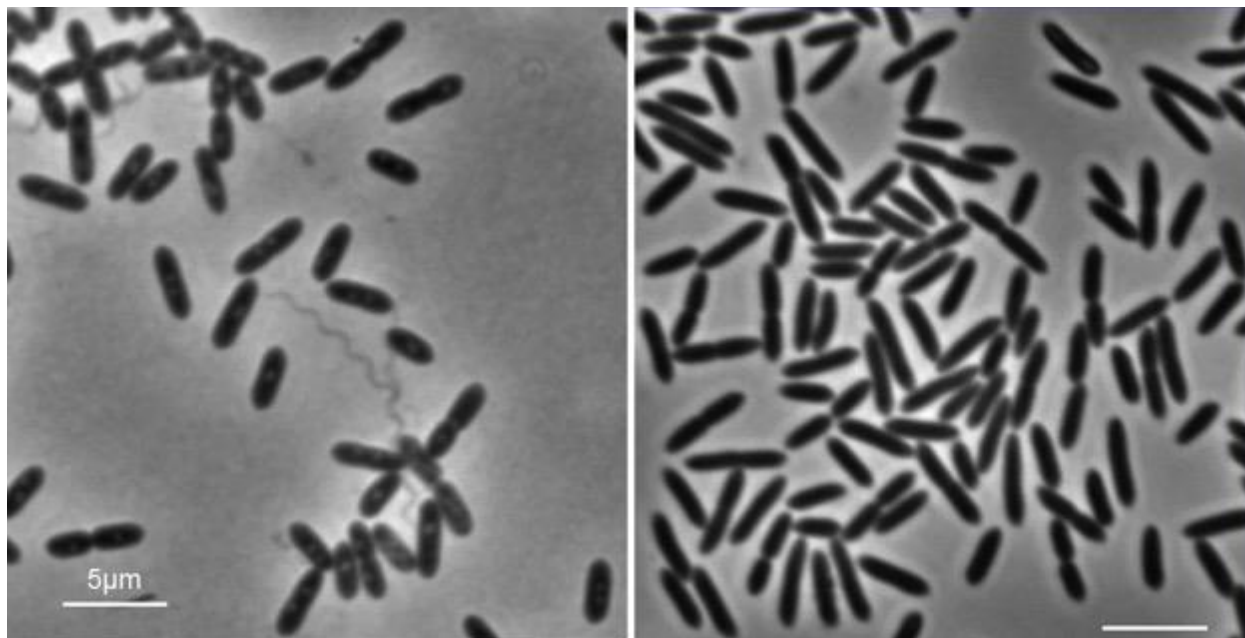


Figure 1. Phase-contrast micrographs of *P. apiarius* NRRL B-23460 and *P. thiaminolyticus* NRRL B-4156. (A) Depicts *P. apiarius* NRRL B-23460 cells grown in TSB for 24 hr at 30°C. (B) Depicts *P. thiaminolyticus* NRRL B-4156 cells grown in TSB for 24 hr at 37°C. Scale bars represent 5 μm .

These paenibacilli were originally classified as members of the genus *Bacillus*, based on their morphological features and biochemical properties, although *P. apiarius*, *P. thiaminolyticus* and their close relatives were not included in the original description of the genus [26]. Due to their similar phenotypes, six strains of *Bacillus thiaminolyticus* were classified in the *B. apiarius* species group, but 16S rRNA gene analysis revealed that *B. apiarius* isolates form two separate clades [25]. This phylogenetic analysis further provided support for reclassifying *B. apiarius* strains as *Paenibacillus apiarius* and those clustering with *B. thiaminolyticus* were renamed [25].

Shortly after, *B. thiaminolyticus* and numerous other *Bacillus* species were reclassified as *Paenibacillus* spp. [27]. Both *P. apiarius* and *P. thiaminolyticus* share the hallmarks of other *Paenibacillus* species in that they are facultative anaerobes, that grow well on nutrient agar at neutral pH but inclusion of a fermentable sugar, such as glucose, will enhance growth [28]. These paenibacilli produce similar colonies when grown for 24 hours on tryptic soy agar and appear circular, entire, and translucent, but are distinguishable by yellow pigmentation of *P. apiarius* colonies, which is not seen with *P. thiaminolyticus*. Both *P. apiarius* and *P. thiaminolyticus* can respire anaerobically using nitrate as an electron acceptor. Both can break down disaccharides and some polysaccharides [27]. Carbon sources that support growth and complex organic compounds that *P. apiarius* and *P. thiaminolyticus* can hydrolyze are listed in Tables 1A and 1B, respectively. Unlike *P. apiarius*, *P. thiaminolyticus* can ferment lactose as well as the sugar alcohols D-mannitol and D-sorbitol [24, 25]. Another distinguishing characteristic is the ability of *P. thiaminolyticus* to produce indole. The ability to decompose thiamin was considered a distinct feature of *P. thiaminolyticus* [24] but can no longer be used to differentiate it from *P. apiarius* or *P. dendritiformis* (unpublished). *P. apiarius* is closely related to the honeybee pathogen *P. alvei*, while *P. thiaminolyticus* is very closely related to *P. dendritiformis*, *P. popilliae*, and *P. lentimorbus*, the latter two species are insect pathogens, responsible for milky spore disease in Japanese beetles [29]. Recently it was discovered that paenibacilli are distinct from *Bacillus* spp. in the arrangement of genes around the chromosomal origin of replication [30]. Paenibacilli code for a YheC/D family protein, designated *orf14*, between the *gyrA* and *gyrB* genes while *Bacillus* species do not have this intervening gene. Our maximum likelihood 16S rRNA gene tree generated by FastTree 2.1 [31] is congruent with these studies (Fig. 2). The tree also indicates that *P. thiaminolyticus* OSY-SE is a strain of *P. apiarius*.

The present study was used to learn more about the genomic context of the thiaminase I gene and thiamin metabolism in these paenibacilli and their close relative *P. dendritiformis* C454 which has a published draft genome [32].

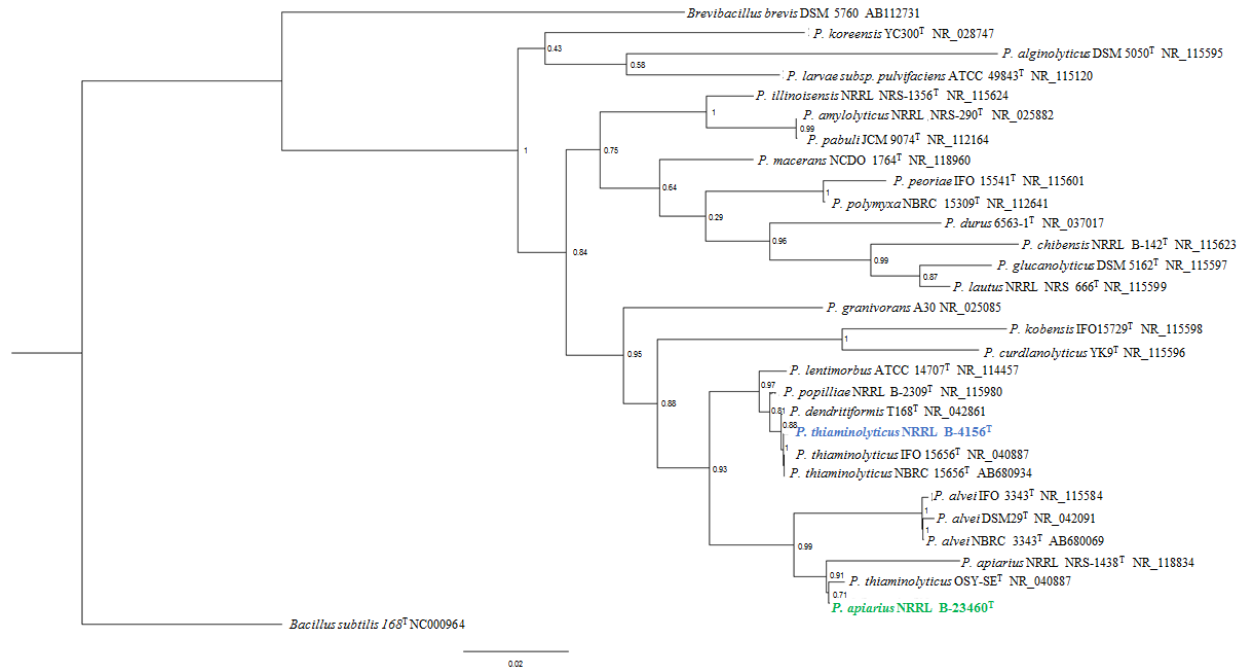


Figure 2. Phylogenetic tree of *Paenibacillus* spp. based on 16S rRNA gene sequences. The maximum likelihood tree was inferred from a comparison of sequences from *Paenibacillus* spp., *Brevibacillus brevis* DSM 5760, and *Bacillus subtilis* 168 using FastTree 2.1 [31]. The sequences generated from the draft genomes of this study are highlighted, with *Paenibacillus thiaminolyticus* NRRL B-4156 in blue font, and *P. apiarius* NRRL B-23460 in green font.

Table 1A. Classification and general features of *P. apiarius* NRRL B-23460 [33]

MIGS ID	Property	Term	Evidence code ^a
	Classification	Domain Bacteria	TAS [55]
		Phylum <i>Firmicutes</i>	TAS [56]
		Class <i>Bacilli</i>	TAS [57, 58]
		Order <i>Bacilliales</i>	TAS [59]
		Family <i>Paenibacillaceae</i>	TAS [57]
		Genus <i>Paenibacillus</i>	TAS [26, 60]
		Species <i>apiarius</i>	TAS [25]
		(Type) strain: <i>NRRL B-23460^T</i>	
	Gram stain	Positive	TAS [25]
	Cell shape	Rod	TAS [25]
	Motility	Motile	TAS [25]
	Sporulation	Endospores with thick coats	TAS [25]
	Temperature range	15-40°C	TAS [25]
	Optimum temperature	28°C	TAS [25]
	pH range; Optimum	Not reported	
	Carbon source	D-glucose, D-galactose, cellobiose, maltose, melibiose, sucrose, trehalose, salicin; can hydrolyze starch, casein	TAS [13]
MIGS-6	Habitat	Soil and honeybee associated	TAS [13]
MIGS-6.3	Salinity	5% NaCl (w/v)	TAS [25]
MIGS-22	Oxygen requirement	facultative	TAS [13]
MIGS-15	Biotic relationship	free-living	TAS [13]
MIGS-14	Pathogenicity	non-pathogen	TAS [25]
MIGS-4	Geographic location	Manitoba, Canada	TAS [13]
MIGS-5	Sample collection	1950s	TAS [13]
MIGS-4.1	Latitude	Not reported	
MIGS-4.2	Longitude	Not reported	
MIGS-4.4	Altitude	Not reported	

^a Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [34]

Table 1B. Classification and general features of *P. thiaminolyticus* NRRL B-4156 [33]

MIGS ID	Property	Term	Evidence code ^a
	Classification	Domain Bacteria	TAS [55]
		Phylum <i>Firmicutes</i>	TAS [56]
		Class <i>Bacilli</i>	TAS [57, 58]
		Order Bacilliales	TAS [59]
		Family -.Paenibacillaceae	TAS [57]
		Genus <i>Paenibacillus</i>	TAS [26, 60]
		Species thiaminolyticus	TAS [24]
		(Type) strain: NRRL B-4156 ^T	
	Gram stain	Positive	TAS [24]
	Cell shape	Rod	TAS [24]
	Motility	Motile	TAS [24]
	Sporulation	endospores	TAS [24]
	Temperature range	20-45°C	TAS [24]
	Optimum temperature	28°C	TAS [24]
	pH range; Optimum	Not reported	
		D-glucose, D-fructose, D-galactose, D-ribose, lactose, cellobiose, maltose, mannose, melibiose, sucrose, trehalose, salicin; can hydrolyze starch, chitin, pullulan, casein	TAS [24]
MIGS-6	Habitat	Soil, animal associated	TAS [24]
MIGS-6.3	Salinity	5% NaCl (w/v)	TAS [24]
MIGS-22	Oxygen requirement	facultative	TAS [24]
MIGS-15	Biotic relationship	free-living	TAS [24]
MIGS-14	Pathogenicity	non-pathogen (1 case in humans)	NAS [21, 24]
MIGS-4	Geographic location	Japan	TAS [24]
MIGS-5	Sample collection	1940s	TAS [24]
MIGS-4.1	Latitude	Not reported	
MIGS-4.2	Longitude	Not reported	
MIGS-4.4	Altitude	Not reported	

^a Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [34]

3.5 GENOME SEQUENCING INFORMATION

Genome project history

Both *P. thiaminolyticus* NRRL B-4156 and *P. apiarius* NRRL B-23460 were acquired from the Agricultural Research Service Culture Collection. The DNA was sequenced in April of 2014. Raw reads were assembled using SPAdes version 5.3 [35]. The contigs were quality filtered by size and coverage. Completeness and heterogeneity were assessed using CheckM [36] and the draft genomes were submitted to Genoscope for annotation with the MicroScope platform [37]. The assembled draft genomes were submitted to the Joint Genome Institute Integrated Microbial Genomes analysis system [38] in October 2016 for annotation. Project summaries are provided in Table 2.

Table 2. Project information.

MIGS ID	Property	Term (<i>P. apiarius</i>)	Term (<i>P. thiaminolyticus</i>)
MIGS 31	Finishing quality	Draft	Draft
MIGS-28	Libraries used	TruSeq	TruSeq
MIGS 29	Sequencing platforms	Illumina MiSeq	Illumina MiSeq
MIGS 31.2	Fold coverage	150x	150x
MIGS 30	Assemblers	SPAdes 5.3	SPAdes 5.3
MIGS 32	Gene calling method	IMG and MicroScope	IMG and MicroScope
	Locus Tag	Ga0138518	Ga0138519
	Genbank ID	NDGJ000000000.1	NDGK000000000.1
	GenBank Date of Release	05/31/2017	05/31/17
	GOLD ID	Ga0138518	Ga0138519
	BIOPROJECT	PRJNA382554	PRJNA382555
MIGS 13	Source Material Identifier	Insect associated	Human associated
	Project relevance	Metabolic pathways	Metabolic pathways

Growth conditions and genomic DNA preparation

Both *P. apiarius* NRRL B-23460 and *P. thiaminolyticus* NRRL B-4156 were grown in tryptic soy broth with shaking, at 30°C and 37°C, respectively. Genomic DNA was extracted using a protocol typically used for isolating high molecular weight DNA from *Bacillus subtilis* [39]. Briefly, cells were lysed with lysozyme and sodium n-lauryl sarcosine. DNA was extracted using phenol:chloroform, and precipitated using ethanol. Near-complete 16S rRNA genes were amplified from the genomic DNA. Sequences were determined and compared with published

sequences available in GenBank. The whole genome sequencing projects for *P. apiarius* NRRL B-23460 and *P. thiaminolyticus* NRRL B-4156 were deposited in DDBJ/EMBL/GenBank under accession numbers NDGJ000000000 and NDGK000000000, respectively.

Genome sequencing and assembly

Illumina MiSeq 2x250 sequencing reactions were conducted on the two DNA samples at the Cornell University Institute of Biotechnology in Ithaca, NY. This resulted in 3,704,766 reads for the *P. apiarius* NRRL B-23460 genome and 4,092,728 reads for the *P. thiamonolyticus* NRRL B-4156 genome. The reads were quality checked and assembled using SPAdes 3.5 [35]. Contigs were filtered based on coverage (above 50x) and size (above 1000 bp). CheckM [36] was used to determine genome completeness and revealed that the *P. apiarius* NRRL B-23460 genome is 99.73% complete with no strain heterogeneity, while the *P. thiaminolyticus* genome is 99.68% complete with no strain heterogeneity.

Genome annotation

Gene calling and annotations for *P. apiarius* NRRL B-23460 and *P. thiaminolyticus* NRRL B-4156 were developed by both the MicroScope platform [37] and IMG [38], *P. dendritiformis* C454 was annotated with MicroScope only. Annotations of interest were independently verified using the Uniprot (Swissprot and TrEMBL) database and BLAST. Ambiguous gene sequences were compared to their *B. subtilis* counterparts to further help identify a putative function. DELTA-BLAST was used to determine functional domains of uncharacterized proteins, and confirm those of characterized proteins of interest.

3.6 GENOME PROPERTIES

The draft genome for *P. apiarius* NRRL B-23460 is 5,404,821 bp (50.49% G+C) and comprises 51 contigs. The largest contig is 827,142 bp, and the smallest is 1,010 bp in length. The N50 of the genome is 280,248. IMG identified 4,957 genes in the genome. Of those genes, 4,822 encode for proteins (97.28%), 22 are rRNA genes (0.44%), 76 are tRNA genes (1.53%), and no pseudogenes were discovered. Of the 22 rRNA genes identified, seven are 5S, ten are 16S, and five are 23S genes. The draft genome for *P. thiaminolyticus* is 6,547,709 bp (53.64% G+C), contains 48 contigs, with the largest contig being 1,172,336 bp and the smallest being 1,148 bp. The N50 is 254,830 bp. For this genome 5,880 genes were identified as protein encoding (97.89%), with 21 rRNA genes (0.36%), 77 tRNA genes (1.31%), and no pseudogenes (0.00%). Amongst the rRNA genes, five 5S, nine 16S, and seven 23S genes were identified. More details of these draft genomes are given in Tables 3A and 3B, and the CoG analyses are summarized in Tables 4A and 4B.

Table 3. Genome and annotation statistics for *P. apiarius* and *P. thiaminolyticus*

Attribute	<i>P. apiarius</i> NRRL B-23460		<i>P. thiaminolyticus</i> NRRL B-4156	
	Value	% of Total	Value	% of Total
Genome size (bp)	5,404,821	100.00	6,537,496	100.00
DNA coding (bp)	4,642,405	85.89	5,508,364	84.26
DNA G+C (bp)	2,729,114	50.49	3,507,168	53.65
DNA scaffolds	51	100.00	47	100.00
Total genes	4,957	100.00	5,880	100.00
Protein coding genes	4,822	97.28	5,756	97.89
RNA genes	135	2.72	124	2.11
Pseudo genes	0	0	0	0
Genes in internal clusters	1,259	25.40	1,709	29.06
Genes with function prediction	3,756	75.77	4,458	75.82
Genes assigned to COGs	3,092	62.38	3,654	62.14
Genes with Pfam domains	3,910	78.88	4,674	79.49
Genes with signal peptides	304	6.13	350	5.95
Genes with transmembrane helices	1,385	27.94	1,658	28.20
CRISPR repeats	0	0	0	0

Table 4A. Number of genes associated with general COG functional categories.

Code	Value	% of total	Description
J	219	6.30%	Translation, ribosomal structure and biogenesis
A	0	0.00%	RNA processing and modification
K	343	9.87%	Transcription
L	98	2.82%	Replication, recombination and repair
B	1	0.03%	Chromatin structure and dynamics
D	50	1.44%	Cell cycle control, Cell division, chromosome partitioning
V	110	3.16%	Defense mechanisms
T	194	5.58%	Signal transduction mechanisms
M	179	5.15%	Cell wall/membrane biogenesis
N	66	1.90%	Cell motility
U	29	0.83%	Intracellular trafficking and secretion
O	112	3.22%	Posttranslational modification, protein turnover, chaperones
C	165	4.75%	Energy production and conversion
G	368	10.59%	Carbohydrate transport and metabolism
E	317	9.12%	Amino acid transport and metabolism
F	103	2.96%	Nucleotide transport and metabolism
H	186	5.35%	Coenzyme transport and metabolism
I	127	3.65%	Lipid transport and metabolism
P	213	6.13%	Inorganic ion transport and metabolism
Q	102	2.93%	Secondary metabolites biosynthesis, transport and catabolism
R	281	8.08%	General function prediction only
S	186	5.35%	Function unknown
-	1,865	37.62%	Not in COGs

The total is based on the total number of protein coding genes in the genome of *P. apiarius* NRRL B-23460.

Table 4B. Number of genes associated with general COG functional categories.

Code	Value	% of total	Description
J	248	6.03%	Translation, ribosomal structure and biogenesis
A	0	0.00%	RNA processing and modification
K	431	10.47%	Transcription
L	112	2.72%	Replication, recombination and repair
B	0	0.00%	Chromatin structure and dynamics
D	59	1.43%	Cell cycle control, Cell division, chromosome partitioning
V	157	3.81%	Defense mechanisms
T	263	6.39%	Signal transduction mechanisms
M	224	5.44%	Cell wall/membrane biogenesis
N	63	1.53%	Cell motility
U	28	0.68%	Intracellular trafficking and secretion
O	142	3.45%	Posttranslational modification, protein turnover, chaperones
C	199	4.83%	Energy production and conversion
G	450	10.93%	Carbohydrate transport and metabolism
E	370	8.99%	Amino acid transport and metabolism
F	109	2.65%	Nucleotide transport and metabolism
H	195	4.74%	Coenzyme transport and metabolism
I	139	3.38%	Lipid transport and metabolism
P	246	5.98%	Inorganic ion transport and metabolism
Q	118	2.87%	Secondary metabolites biosynthesis, transport and catabolism
R	334	8.11%	General function prediction only
S	202	4.91%	Function unknown
-	2,226	37.86%	Not in COGs

The total is based on the total number of protein coding genes in the genome of *P. thiaminolyticus* NRRL B-4156

3.7 INSIGHTS FROM THE GENOME

We investigated the potential thiamin biosynthetic capabilities of *P. apiarius* NRRL B-23460, *P. thiaminolyticus* NRRL B-4156, and *P. dendritiformis* C454 using the annotations and metabolic pathways generated by MicroScope. Typically in bacteria, TPP, the active cofactor, is formed from two phosphorylated moieties, THZ-P and HMP-PP. The thiazole moiety is derived from the glycolysis products pyruvate and G3P, a sulfur from cysteine, and either tyrosine (in *E. coli*) or glycine (in *B. subtilis*) [40]. The formation of THZ-P requires a suite of proteins including

deoxy-d-xylulose 5-phosphate synthase (Dxs), a sulfur donor protein (NifS or IscS), adenylyltransferase (ThiF), sulfur carrier protein (ThiS), thiazole synthase (ThiG), thiazole biosynthesis protein ThiH or glycine oxidase ThiO, and in some cases an aromatase (TenI) [40]. The pyrimidine moiety is derived from AIR, an intermediate in purine biosynthesis. HMP synthase (ThiC) and HMP kinase (ThiD) are required to form HMP-PP [40]. Thiamin phosphate synthase (ThiE) combines THZ-P and HMP-PP to form TMP, which is then phosphorylated by thiamin phosphate kinase (ThiL), forming TPP [40].

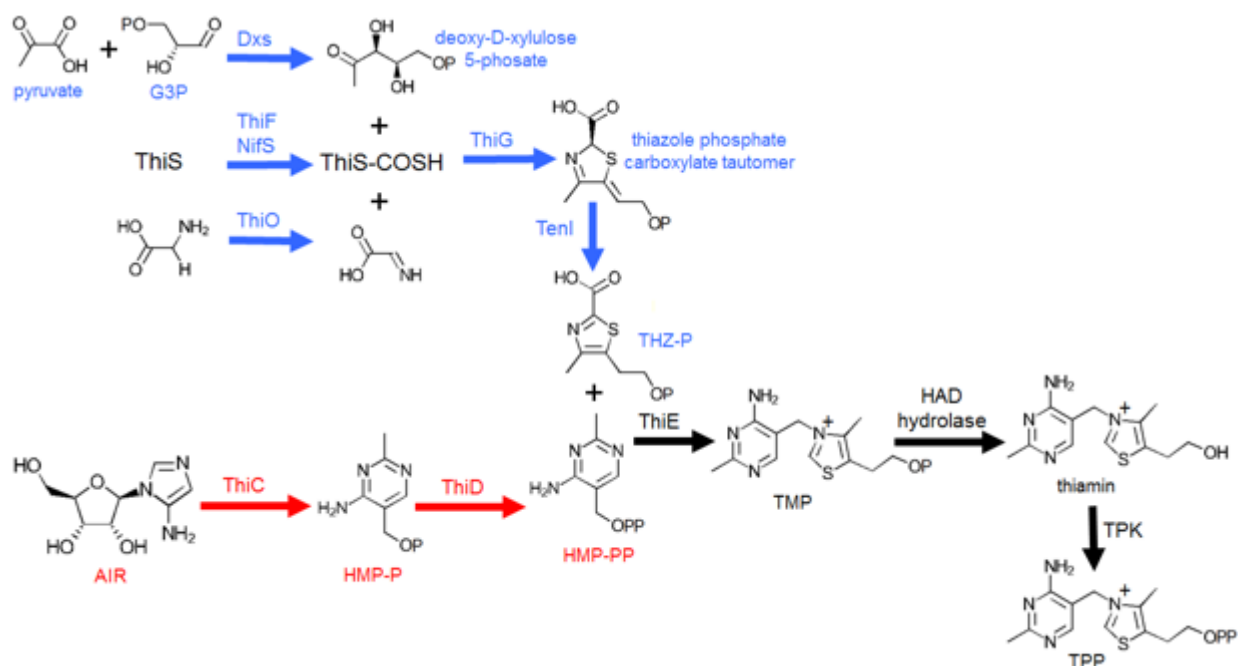


Figure 3. Predicted thiamin biosynthesis pathway in *P. apiarius* NRRL B-23460. Pathways involved in thiazole biosynthesis are highlighted in blue and pathways involved in pyrimidine biosynthesis are shown in red. The steps in black correspond to the coupling of the thiazole and pyrimidine moieties, and the formation of the active cofactor TPP.

The *P. apiarius* draft genome contains all of the genes required to make THZ-P (Fig. 3); *thiG*, *thiO*, *thiS*, and *tenI* are located in an operon putatively regulated by a TPP-binding riboswitch. The presence of *thiO* suggests that NRRL B-23460 uses glycine to generate the thiazole moiety. With *thiC* and *thiD* present, it appears competent for HMP-PP biosynthesis (3).

The genome contains *thiE* but lacks *thiL* for the terminal phosphorylation; however, it contains a putative TPK. Plants, fungi, and a few species of bacteria, use a different thiamin biosynthesis strategy where thiamin monophosphate is dephosphorylated to thiamin, and then converted to TPP by a TPK [41]. The genome content of *P. apiarius* NRRL B-23460 suggests that it synthesizes thiamin in this manner. Hasnain and colleagues recently demonstrated that HAD-superfamily enzymes of the subfamily IA in plants and some bacteria catalyze the dephosphorylation of TMP [41]. In bacteria, these hydrolase genes are either fused to a thiamin biosynthesis gene, like *thiD* or *thiE*, or these genes are located in the same operon. The *P. apiarius* NRRL B-23460 genome has a HAD subfamily IA hydrolase gene that potentially serves this function, and is located in an operon with other thiamin biosynthesis genes (Fig. 4).



Figure 4. Putative thiaminase I operons in the three *Paenibacillus* species. Annotations of conserved genes shared amongst the three species are as follows (1) *thiM*; (2) *thiD*; (3) *thiE*; (4) glycosyltransferase; (5) thymidylate synthase; (6) methyltransferase; (7) *yzgD* NUDIX hydrolase; (8) thiaminase I and (9) HAD hydrolase. Three genes in *P. apiarius* B23460 that are not conserved include a putative transcriptional regulator and two genes of unknown function. The operons may be under the control of a putative thiamin pyrophosphate binding riboswitch, upstream of *thiM*. Genes are color-coded based on proposed functions.

The genes *thiD*, *thiE*, and HAD subfamily IA hydrolase gene are found in a highly conserved operon with the thiaminase I gene in the *P. apiarius* NRRL B-23460, *P. thiaminolyticus* NRRL B-4156 and *P. dendritiformis* C454 draft genomes. The operons of all three strains are depicted in Figure 4 and appear to be regulated by a TPP-binding riboswitch [42]. Thiazole kinase (*thiM*), which phosphorylates THZ (Fig. 5) [40] is the first gene in the operon, followed by *thiD* and *thiE*. The thiamin biosynthesis genes are preceded by a nucleoside 2-deoxyribosyltransferase,

a thymidylate synthase, a SAM-dependent methyltransferase, a Nudix-family hydrolase (YzgD), and thiaminase I (Fig. 4). In the *P. thiaminolyticus* NRRL B-4156 and *P. dendritiformis* C454 operons, a HAD subfamily IA hydrolase is located directly after the thiaminase I gene. In *P. apiarius* NRRL B-23460, three additional genes are present, which code for a putative transcriptional regulator and two proteins of unknown function (Fig. 4). Since the HAD hydrolase is in the same operon as *thiD* and *thiE*, it is likely that it performs the dephosphorylation of TMP. Biochemical studies by Tirrell and colleagues reveal that the YzgD Nudix hydrolase has a HAD domain, which specifically cleaves pyridoxal phosphate, but does not dephosphorylate TMP, TPP, or THZ-P although HMP-P was not tested [43]. The Nudix hydrolase domain is more promiscuous as it is active on nucleoside diphosphates such as CDP-alcohols, ADP-coenzymes, ADP-ribose, TDP-glucose, and some UDP-sugars, restoring the nucleoside monophosphate [43]. It is unclear how this enzyme relates to thiaminase I, but it may play a role in thiamin metabolism. Recently, Nudix hydrolases were discovered clustered with thiamin biosynthesis genes in a few bacterial species as well as in plants and yeast. These Nudix proteins are able to hydrolyze a phosphate from the diphosphate forms of oxothiamin and oxythiamin, thiamin oxidation and hydrolysis products respectively, providing these cells with resistance to these toxic analogs [44]. Due to its location in the paenibacilli genomes, it may serving this protective function, preventing the cell from using toxic thiamin analogs as cofactors instead of TPP.

Cooper et al. described the bacimethrin operon of *Clostridium botulinum*, which includes the thiaminase I gene [45]. Bacimethrin is a toxic analog of HMP, that when combined with THZ-P forms the antivitamin 2'-methoxythiamin pyrophosphate, which binds enzymes in place of the TPP cofactor, thus rendering the enzyme nonfunctional [45-47]. The bacimethrin operon consists of a glycosyltransferase (nucleoside 2-deoxyribosyltransferase), thymidylate synthase,

methyltransferase, thiaminase I, and pyrimidine kinase, all of which are present in the three paenibacilli (Fig. 4), making it likely that they can produce bacimethrin [45]. In the paenibacilli, ThiD may be bifunctional, serving as a kinase for both pyrimidines, phosphorylating bacimethrin as well as HMP-P. The function of thiaminase I when the antivitamin is produced is not known. Since thiaminase I does not degrade 2'-methoxythiamin pyrophosphate in *C. botulinum* [45], it is possible that thiaminase I could enhance the effectiveness of this antibiotic against competing bacteria. In contrast to the paenibacilli operons, the *C. botulinum* thiaminase I operon does not contain genes involved in thiamin biosynthesis and salvage. The *C. botulinum* operon also contains a putative ABC transporter that is not found in the paenibacilli thiaminase I operons.

Apparently, *P. dendritiformis* C454 and *P. thiaminolyticus* NRRL B-4156 lack the genomic potential to synthesize both moieties of thiamin. Of the genes involved in thiazole biosynthesis, they both have *dxs* and *nifS*, and NRRL B-4156 contains *thiF*. Neither has *thiO* which is essential for thiazole synthesis in *B. subtilis*. Both lack *thiC*, so they are unable to convert AIR to HMP. The presence of *thiD* and *thiE* in their thiaminase I operons provides the potential to make TMP from environmentally acquired THZ and HMP, a strategy used by other bacteria [48]. ThiM can phosphorylate environmentally derived thiazole alcohol, which can be combined with HMP-P by ThiE (Fig. 5). Like *P. apiarius* NRRL B-23460, their genomes encode TPK to make TPP. The presence of the thiaminase I in the same operon as *thiM*, *thiD*, and *thiE* suggests a potential role in thiamin salvage. The thiaminase I, acting on thiamin or pyrithiamine (a thiamin analog) [10] would generate HMP* and a free THZ. We propose that THZ and HMP* could be imported into the cell, phosphorylated by their respective kinases, and combined by ThiE (Fig. 5). TPP is then produced via dephosphorylation by the HAD hydrolase and addition of the pyrophosphate by TPK.

The potential pathways available to the three paenibacilli to salvage thiamin are summarized in Figure 5. Both *P. dendritiformis* C454 and *P. thiaminolyticus* NRRL B-4156 genomes code for the intracellular enzyme thiaminase II (TenA), but *P. apiarius* NRRL B-23460 lacks this gene. Thiaminase II catalyzes the base exchange of thiamin with water, but is not a thiaminase I homolog [49]. It functions in the salvage of HMP from base-degraded thiamin [49]. In both genomes that code for this enzyme, TenA appears regulated by a TPP riboswitch. The genomes of all three paenibacilli contain *ylmB*, which deacetylates base-degraded thiamin forming aminopyrimidine, the preferred substrate for TenA [49] (Fig. 5).

MicroScope identified another TPP riboswitch in all three genomes that appears to regulate a transport system. In all three operons, the riboswitch is preceded by an NMT1/Thi5 domain protein. Thi5 is a yeast protein that converts pyridoxal and histidine to HMP-P, and is a homolog to the ThiY protein found in *Bacillus cereus* and *B. halodurans* [50]. ThiY is part of the ThiXYZ ABC transport system putatively involved in the uptake of HMP, as well as in the uptake of base-degraded thiamin [49-51]. In *P. apiarius* NRRL B-23460, this ThiY homolog is followed by a small, 98 amino acid protein with a thiamin-binding domain, suggesting it may have two alternative transporters for this system. However, in the other two paenibacilli genomes, this is followed by a permease, and the ATP-binding protein of the ABC transport system. The other two ABC transport proteins are found after the small thiamin-binding protein in NRRL B-23460. This suggests that the genomes of all three paenibacilli contain the ThiXYZ HMP transport system, or a homologous system. In *P. thiaminolyticus* NRRL B-4156 and *P. dendritiformis* C454, the system could potentially be dedicated for base-degraded thiamin uptake. The lack of TenA in *P. apiarius* NRRL B-23460 may explain why it has an additional thiamin-binding protein associated with this

transport system, as it cannot use base-degraded thiamin. It is also plausible that this system is involved in the uptake of the HMP* generated by thiaminase I in all three species.

All three paenibacilli contain the *ykoC-F* operon, which encodes for a putative ABC transport system for HMP uptake [52]. The genes encode for two transmembrane components, an ATPase, and an HMP/thiamin-binding protein YkoF [52]. It is unclear if this system takes up both HMP and thiamin, or is specific for HMP and HMP derivatives, as YkoF binds the HMP moiety and does not appear to have any residues to anchor the thiazole moiety of thiamin. This is in contrast to thiamin binding by TbpA, which also binds the THZ [12]. The YkoF transporter could potentially be used for the uptake of the HMP* derived from thiaminase I breakdown of thiamin and thiamin analogs, as well as free HMP, and possibly base-degraded thiamin as well. MicroScope identified a TPP-binding riboswitch upstream of this operon in all three paenibacilli genomes, suggesting that its expression is regulated by thiamin availability.

The *P. dendritiformis* C454 and *P. thiaminolyticus* NRRL B-4156 genomes contain another thiamin ABC transport permease in addition to the YkoC-F system. The thiamin permeases in these two genomes appear to be regulated by TPP riboswitches and share amino acid sequence similarity with YkoD. Next to the permease is the ATP-binding protein, and the third gene in the operon encodes another transmembrane permease with homology to the cobalt ABC transporter permease CbiQ. The presence of this permease suggests that the *yko* system is only used in HMP and HMP derivative uptake and this system is specific for thiamin, allowing for the two thiamin auxotrophs to acquire intact thiamin from the environment. In all three genomes, ThiW, a transporter specific for THZ [53, 54], was not identified. It is possibly they acquire environmental THZ via an unknown mechanism.

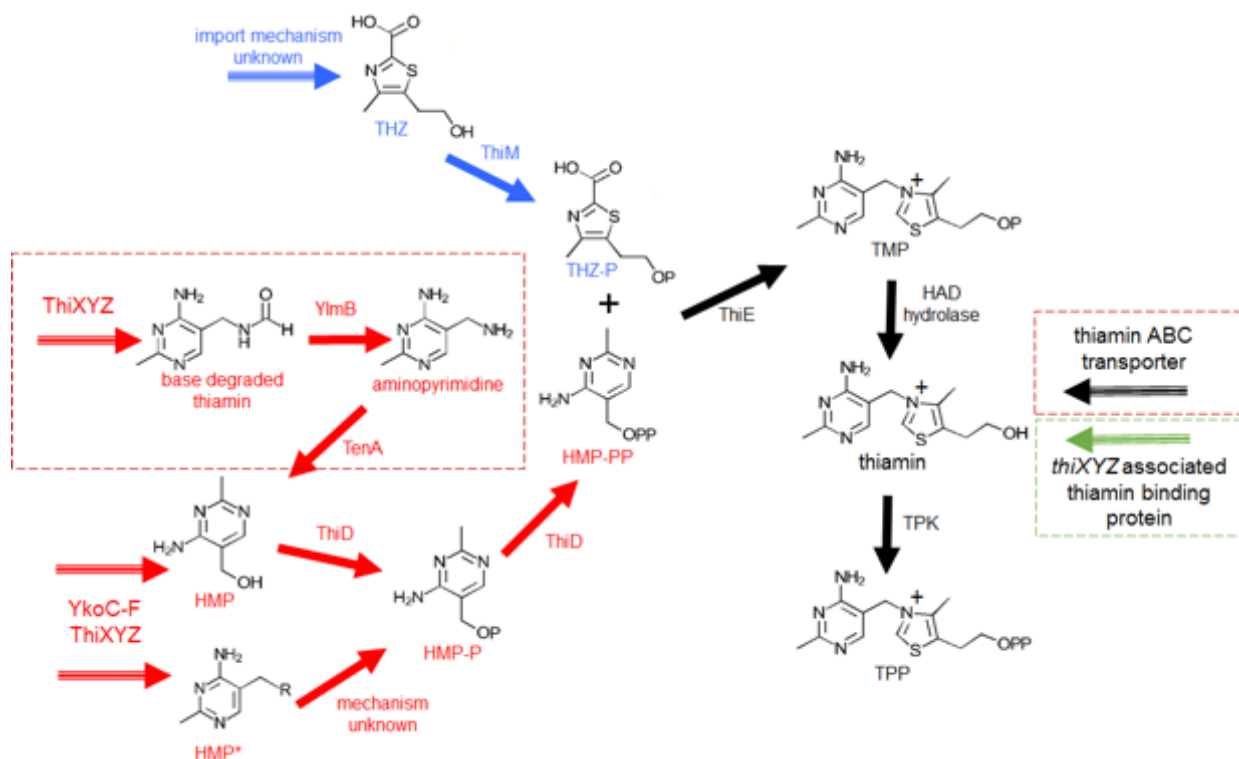


Figure 5. Predicted thiamin salvage pathways in all three *paenibacilli*. Pathways involved in thiazole salvage are highlighted in blue and pathways involved in pyrimidine salvage are shown in red. The dotted red boxes are steps unique to *P. thiaminolyticus* NRRL B-4156 and *P. dendritiformis* C454. Biosynthetic pathways are shown with solid arrows and import pathways are indicated with a striped arrow. The putative importer boxed in green is unique to *P. apiarius* NRRL B-23460. In all cases, it is not yet understood how THZ enters the cell.

3.8 CONCLUSIONS

The genome sequences of *P. apiarius* NRRL B-23460 and *P. thiaminolyticus* NRRL B-4156 reveal insights into thiamin metabolism of these organisms. While *P. apiarius* NRRL B-23460 appears capable of synthesizing thiamin *de novo*, *P. thiaminolyticus* NRRL B-4156 is not, as it lacks the ability to make HMP and THZ. Both organisms apparently phosphorylate thiamin to its active form in a manner rarely used in bacteria, as they can dephosphorylate TMP and then add two phosphates with a pyrophosphokinase to make TPP. The thiaminase I gene is located in a putatively TPP riboswitch-regulated operon with genes for the synthesis of bacimethrin, as well as thiamin biosynthesis and salvage genes. This suggests a potential metabolic role for thiaminase I

in thiamin synthesis, especially in *P. thiaminolyticus* NRRL B-4156, which cannot synthesize thiamin precursors. Further, both species appear to have two different systems to take up HMP, both of which appear to be regulated with TPP riboswitches. It is possible that one of these transport systems is specific for HMP* generated from thiaminase I. We suggest that this HMP* compound can be used in thiamin biosynthesis along with THZ scavenged from the breakdown of thiamin. *P. thiaminolyticus* NRRL B-4156 has the ability to salvage base-degraded thiamin with its intracellular thiaminase II [49], whereas *P. apiarius* NRRL B-23460 does not. This is another method in which *P. thiaminolyticus* can acquire the pyrimidine precursor for thiamin. To further compensate for its auxotrophy, *P. thiaminolyticus* NRRL B-4156 may have a thiamin specific ABC transport system, which is not present in the *P. apiarius* NRRL B-23460 genome. However, NRRL B-23460 has a unique thiamin-binding protein encoded for in the *thiXYZ* operon which *P. thiaminolyticus* NRRL B-4156 lacks. Biochemical and genetic tests need to be conducted to test the hypotheses generated in this study to further elucidate the roles these genes and proteins play in thiamin metabolism.

3.9 REFERENCES

1. Schumpeter EB: **The Industrialization of Japan and Manchukuo, 1930-1940**, vol. 8: Taylor & Francis US; 1940.
2. Fujita A: **Thiaminase**. *Advances in Enzymology and Related Areas of Molecular Biology, Volume 15* 1954:389-421.
3. Fujita A, Nose Y, Kozuka S, Tashiro T, Ueda K, Sakamoto S: **Studies on thiaminase. 1. Activation of thiamine breakdown by organic bases**. *Journal of Biological Chemistry* 1952, **196**:289-295.
4. Kuno Y: **Bacillus thiaminolyticus, a new thiamin-decomposing bacterium**. *Proceedings of the Japan Academy* 1951, **27**(7):362-365.
5. Abe M, Ito S-i, Kimoto M, Hayashi R, Nishimune T: **Molecular studies on thiaminase I**. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression* 1987, **909**(3):213-221.
6. Douthit H, Airth R: **Thiaminase I of Bacillus thiaminolyticus**. *Archives of biochemistry and biophysics* 1966, **113**(2):331-337.

7. Costello CA, Kelleher NL, Abe M, McLafferty FW, Begley TP: **Mechanistic studies on thiaminase I overexpression and identification of the active site nucleophile.** *Journal of Biological Chemistry* 1996, **271**(7):3445-3452.
8. Wang L, Wilkins JH, Airth R: **Repression of thiaminase I by thiamine and related compounds in *Bacillus thiaminolyticus*.** *Canadian journal of microbiology* 1968, **14**(10):1143-1147.
9. Wang L, Airth R: **Repression of thiaminase I in *Bacillus thiaminolyticus*.** *Biochemical and biophysical research communications* 1967, **27**(3):325-330.
10. Agee CC, Airth R: **Reversible inactivation of thiaminase I of *Bacillus thiaminolyticus* by its primary substrate, thiamine.** *Journal of bacteriology* 1973, **115**(3):957-965.
11. Ebata J, Murata K: **The purification of thiaminase I produced by *Bacillus thiaminolyticus*.** *The Journal of vitaminology* 1961, **7**(2):115-121.
12. Soriano EV, Rajashankar KR, Hanes JW, Bale S, Begley TP, Ealick SE: **Structural similarities between thiamin-binding protein and thiaminase-I suggest a common ancestor.** *Biochemistry* 2008, **47**(5):1346-1357.
13. Katznelson H: ***Bacillus apiarius*, n. sp., an aerobic spore-forming organism isolated from honeybee larvae.** *Journal of bacteriology* 1955, **70**(6):635.
14. Kraft CE, Angert ER: **Competition for vitamin B1 (thiamin) structures numerous ecological interactions.** *The Quarterly Review of Biology* 2017, **92**(2):151-168.
15. Kraft CE, Gordon ER, Angert ER: **A Rapid Method for Assaying Thiaminase I Activity in Diverse Biological Samples.** *PloS one* 2014, **9**(3):e92688.
16. McCleary BV, Chick BF: **The purification and properties of a thiaminase I enzyme from *Nardoo* (*Marsilea drummondii*).** *Phytochemistry* 1977, **16**(2):207-213.
17. Kreinbring CA, Remillard SP, Hubbard P, Brodtkin HR, Leeper FJ, Hawksley D, Lai EY, Fulton C, Petsko GA, Ringe D: **Structure of a eukaryotic thiaminase I.** *Proceedings of the National Academy of Sciences* 2014, **111**(1):137-142.
18. Honeyfield DC, Hinterkopf JP, Brown SB: **Isolation of thiaminase-positive bacteria from alewife.** *Transactions of the American Fisheries Society* 2002, **131**(1):171-175.
19. Honeyfield DC, Hinterkopf JP, Fitzsimons JD, Tillitt DE, Zajicek JL, Brown SB: **Development of thiamine deficiencies and early mortality syndrome in lake trout by feeding experimental and feral fish diets containing thiaminase.** *J Aquat Anim Health* 2005, **17**(1):4-12.
20. Richter CA, Evans AN, Wright-Osment MK, Zajicek JL, Heppell SA, Riley SC, Krueger CC, Tillitt DE: ***Paenibacillus thiaminolyticus* is not the cause of thiamine deficiency impeding lake trout (*Salvelinus namaycush*) recruitment in the Great Lakes.** *Canadian Journal of Fisheries and Aquatic Sciences* 2012, **69**(6):1056-1064.
21. Ouyang J, Pei Z, Lutwick L, Dalal S, Yang L, Cassai N, Sandhu K, Hanna B, Wiczorek RL, Bluth M: ***Paenibacillus thiaminolyticus*: a new cause of human infection, inducing bacteremia in a patient on hemodialysis.** *Annals of Clinical & Laboratory Science* 2008, **38**(4):393-400.
22. Thomas K, Griffiths F: **Natural establishment of thiaminase activity in the alimentary tract of newborn lambs and effects on thiamine status and growth rates.** *Australian veterinary journal* 1987, **64**(7):207-210.
23. Richter CA, Wright-Osment MK, Zajicek JL, Honeyfield DC, Tillitt DE: **Quantitative polymerase chain reaction (PCR) assays for a bacterial thiaminase I gene and the**

- thiaminase-producing bacterium *Paenibacillus thiaminolyticus***. *Journal of aquatic animal health* 2009, **21**(4):229-238.
24. Nakamura L: ***Bacillus thiaminolyticus* sp. nov., nom. rev.** *International Journal of Systematic and Evolutionary Microbiology* 1990, **40**(3):242-246.
 25. Nakamura L: ***Paenibacillus apiarius* sp. nov.** *International Journal of Systematic and Evolutionary Microbiology* 1996, **46**(3):688-693.
 26. Ash C, Priest FG, Collins MD: **Molecular identification of rRNA group 3 bacilli (Ash, Farrow, Wallbanks and Collins) using a PCR probe test.** *Antonie van Leeuwenhoek* 1993, **64**(3-4):253-260.
 27. Shida O, Takagi H, Kadowaki K, Nakamura LK, Komagata K: **Transfer of *Bacillus alginolyticus*, *Bacillus chondroitinus*, *Bacillus curdlanolyticus*, *Bacillus glucanolyticus*, *Bacillus kobensis*, and *Bacillus thiaminolyticus* to the genus *Paenibacillus* and emended description of the genus *Paenibacillus*.** *International Journal of Systematic and Evolutionary Microbiology* 1997, **47**(2):289-298.
 28. Schleifer K-H: **Phylum XIII. Firmicutes Gibbons and Murray 1978, 5 (Firmacutes [sic] Gibbons and Murray 1978, 5).** In: *Bergey's Manual® of Systematic Bacteriology: Volume Three The Firmicutes*. Edited by De Vos P, Garrity GM, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer K-H, Whitman WB. New York, NY: Springer New York; 2009: 19-1317.
 29. Rippere KE, Tran MT, Yousten AA, Hilu KH, Klein MG: ***Bacillus popilliae* and *Bacillus lentimorbus*, bacteria causing milky disease in Japanese beetles and related scarab larvae.** *International Journal of Systematic and Evolutionary Microbiology* 1998, **48**(2):395-402.
 30. Iiyama K, Otao M, Mori K, Mon H, Lee JM, Kusakabe T, Tashiro K, Asano S-I, Yasunaga-Aoki C: **Phylogenetic relationship of *Paenibacillus* species based on putative replication origin regions and analysis of an yheCD-like sequence found in this region.** *Bioscience, biotechnology, and biochemistry* 2014, **78**(5):891-897.
 31. Price MN, Dehal PS, Arkin AP: **FastTree 2—approximately maximum-likelihood trees for large alignments.** *PloS one* 2010, **5**(3):e9490.
 32. Sirota-Madi A, Olender T, Helman Y, Brainis I, Finkelshtein A, Roth D, Hagai E, Leshkowitz D, Brodsky L, Galatenko V: **Genome sequence of the pattern-forming social bacterium *Paenibacillus dendritiformis* C454 chiral morphotype.** *Journal of bacteriology* 2012, **194**(8):2127-2128.
 33. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV: **The minimum information about a genome sequence (MIGS) specification.** *Nature biotechnology* 2008, **26**(5):541-547.
 34. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT: **Gene Ontology: tool for the unification of biology.** *Nature genetics* 2000, **25**(1):25-29.
 35. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD: **SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing.** *Journal of computational biology* 2012, **19**(5):455-477.
 36. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW: **CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes.** *Genome research* 2015, **25**(7):1043-1055.

37. Vallenet D, Belda E, Calteau A, Cruveiller S, Engelen S, Lajus A, Le Fèvre F, Longin C, Mornico D, Roche D: **MicroScope—an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data.** *Nucleic acids research* 2012;gks1194.
38. Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P: **IMG: the integrated microbial genomes database and comparative analysis system.** *Nucleic acids research* 2012, **40**(D1):D115-D122.
39. Harwood CR, Cutting SM: **Molecular biological methods for Bacillus.** Chichester; New York: Wiley; 1990.
40. Jurgenson CT, Begley TP, Ealick SE: **The structural and biochemical foundations of thiamin biosynthesis.** *Annu Rev Biochem* 2009, **78**:569-603.
41. Hasnain G, Roje S, Sa N, Zallot R, Ziemak MJ, de Crécy-Lagard V, Gregory JF, Hanson AD: **Bacterial and plant HAD enzymes catalyse a missing phosphatase step in thiamin diphosphate biosynthesis.** *Biochemical Journal* 2016, **473**(2):157-166.
42. Mironov AS, Gusarov I, Rafikov R, Lopez LE, Shatalin K, Kreneva RA, Perumov DA, Nudler E: **Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria.** *Cell* 2002, **111**(5):747-756.
43. Tirrell IM, Wall JL, Daley CJ, Denial SJ, Tennis FG, Galens KG, O'Handley SF: **YZGD from Paenibacillus thiaminolyticus, a pyridoxal phosphatase of the HAD (haloacid dehalogenase) superfamily and a versatile member of the Nudix (nucleoside diphosphate x) hydrolase superfamily.** *Biochemical Journal* 2006, **394**(3):665-674.
44. Goyer A, Hasnain G, Frelin O, Ralat MA, Gregory JF, Hanson AD: **A cross-kingdom Nudix enzyme that pre-empts damage in thiamin metabolism.** *Biochemical Journal* 2013, **454**(3):533-542.
45. Cooper LE, O'Leary SnE, Begley TP: **Biosynthesis of a Thiamin Antivitamin in Clostridium botulinum.** *Biochemistry* 2014, **53**(14):2215-2217.
46. Reddick JJ, Saha S, Lee J-m, Melnick JS, Perkins J, Begley TP: **The mechanism of action of bacimethrin, a naturally occurring thiamin antimetabolite.** *Bioorganic & medicinal chemistry letters* 2001, **11**(17):2245-2248.
47. Zilles JL, Croal LR, Downs DM: **Action of the thiamine antagonist bacimethrin on thiamine biosynthesis.** *Journal of bacteriology* 2000, **182**(19):5606-5610.
48. Karunakaran R, Ebert K, Harvey S, Leonard M, Ramachandran V, Poole P: **Thiamine is synthesized by a salvage pathway in Rhizobium leguminosarum bv. viciae strain 3841.** *Journal of bacteriology* 2006, **188**(18):6661-6668.
49. Jenkins AH, Schyns G, Potot S, Sun G, Begley TP: **A new thiamin salvage pathway.** *Nature chemical biology* 2007, **3**(8):492-497.
50. Bale S, Rajashankar KR, Perry K, Begley TP, Ealick SE: **HMP binding protein ThiY and HMP-P synthase THI5 are structural homologues.** *Biochemistry* 2010, **49**(41):8929-8936.
51. Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS: **Comparative genomics of thiamin biosynthesis in procaryotes New genes and regulatory mechanisms.** *Journal of Biological chemistry* 2002, **277**(50):48949-48959.
52. Devedjiev Y, Surendranath Y, Derewenda U, Gabrys A, Cooper DR, Zhang R-g, Lezondra L, Joachimiak A, Derewenda ZS: **The structure and ligand binding properties of the B. subtilis YkoF gene product, a member of a novel family of thiamin/HMP-binding proteins.** *Journal of molecular biology* 2004, **343**(2):395-406.

53. Rodionov DA, Hebbeln P, Eudes A, Ter Beek J, Rodionova IA, Erkens GB, Slotboom DJ, Gelfand MS, Osterman AL, Hanson AD: **A novel class of modular transporters for vitamins in prokaryotes.** *Journal of bacteriology* 2009, **191**(1):42-51.
54. Anderson LN, Koech PK, Plymale AE, Landorf EV, Konopka A, Collart FR, Lipton MS, Romine MF, Wright AT: **Live cell discovery of microbial vitamin transport and enzyme-cofactor interactions.** *ACS chemical biology* 2015, **11**(2):345-354.

CHAPTER 4

A ROLE FOR THIAMINASE I IN THIAMIN SALVAGE[†]

4.1 ABSTRACT

Thiamin is essential for life, as it serves as a cofactor for enzymes involved in critical carbon transformations. It is composed of two moieties, which are combined and phosphorylated to form the active cofactor. Thiamin can be synthesized by many bacteria, while thiamin auxotrophs must obtain thiamin or its moieties from the environment. Thiaminases are enzymes that degrade thiamin by catalyzing the base-exchange substitution of the thiazole with a nucleophile. Thiaminase II is involved in the salvage of base-degraded thiamin, whereas the function of the non-homologous thiaminase I is a mystery. Unlike thiaminase II, thiaminase I is an extracellular enzyme only known to be produced by a few bacteria. Thiaminase I activity has been associated with plants and animals, and it has been implicated in thiamin deficiency syndromes in animals. We used the thiaminase I producing soil bacterium *Burkholderia thailandensis* as a genetically tractable model to ascertain the function of thiaminase I. We found that thiaminase I extended the survival of the thiamin auxotrophs we generated, when grown in defined media when compared to isogenic strains that could not produce thiaminase I. These results were confirmed by the addition of purified thiaminase I to cultures. We also found that *B. thailandensis* thiamin auxotrophs lacking thiaminase I prefer to grow on thiamin precursors rather than thiamin itself. Our findings demonstrate that thiaminase I functions in thiamin salvage by converting thiamin and thiamin analogs to their precursor moieties which can then be used for growth. These results provide the first evidence for a biologically relevant benefit of thiaminase I toward a thiaminase I producing bacterium.

[†] results from this chapter are to be submitted to the Journal of Bacteriology and it is written according to their guidelines

4.2 IMPORTANCE

The function of thiaminase I has remained a long standing, unsolved mystery. It is only known to be produced by a small subset of microorganisms although thiaminase I activity has been associated with numerous plants and animals, and is implicated in thiamin deficiencies brought on by consumption of the enzyme. Genomic and biochemical analysis have shed light on potential roles for the enzyme. Using the genetically amenable thiaminase I producing soil bacterium *Burkholderia thailandensis*, we were able to demonstrate thiaminase I helps salvage precursors from thiamin and thiamin analogs in the environment. Our study is the first to establish a biological role for this perplexing enzyme and provides insight into the complicated nature of thiamin metabolism. It also establishes *B. thailandensis* as a robust model system for studying thiamin metabolism.

4.3 INTRODUCTION

Thiamin (vitamin B₁) is an essential vitamin necessary for all cellular life, with only one known exception (1). Its diphosphorylated form, thiamin pyrophosphate (TPP), serves as a cofactor for enzymes such as transketolase and α -ketoglutarate dehydrogenase that perform critical carbon transformations necessary for energy metabolism and the biosynthesis of precursors for cellular macromolecules (2). Thiamin can be synthesized by plants, fungi and numerous bacteria. The vitamin is composed of pyrimidine and thiazole (THZ) moieties, which are synthesized separately and then combined in biological systems. Typically, in bacteria, the pyrimidine moiety is derived from the purine intermediate 5-aminoimidazole ribotide. Hydroxymethyl pyrimidine (HMP) synthase, ThiC, converts it to HMP-P, which is then phosphorylated to HMP-PP by HMP kinase (ThiD). The THZ moiety is synthesized from glyceraldehyde 3-phosphate, pyruvate, the

sulfur carrier protein ThiS, and glycine or tyrosine in a multistep process where thiazole synthase (ThiG) ultimately combines the precursors to form THZ-P carboxylate (3). Thiamin phosphate synthase (ThiE) combines THZ-P and HMP-PP to form thiamin monophosphate (TMP). Typically in bacteria this compound is directly phosphorylated to form TPP by ThiL, however, some bacteria dephosphorylate TMP to thiamin and then add a pyrophosphate, as synthesized in plants and fungi (refer to (3) for more detail on thiamin biosynthesis).

Although the ability to synthesize thiamin is common in bacteria, many species are thiamin auxotrophs. *Listeria monocytogenes* abrogates the need to synthesize thiamin through uptake via the ThiT thiamin specific transporter (4). Members of the Enterobacteriaceae use the ThiBPQ system, which not only imports thiamin but its phosphorylated forms as well (5). Imported intracellular thiamin can be diphosphorylated directly to TPP via thiamin pyrophosphokinase or stepwise to TMP by thiamin kinase and then to TPP by ThiL (6). Some bacteria are able to salvage thiamin precursors for thiamin synthesis (7, 8). This is achieved through transporters specific for HMP, such as the YkoCDEF system in *Bacillus subtilis* (9, 10), the CytX permease (11, 12), and HmpT (10). The import of THZ is less well defined in bacteria, but a THZ specific transporter, ThiW, has been described in multiple species (10, 12). After import, the precursors are processed through the later stages of the synthetic pathway to form an active cofactor.

Bacteria also produce thiaminases, enzymes that degrade thiamin to its constituents by catalyzing the base exchange of thiazole with a nucleophile. Thiaminases are categorized in two classes, thiaminase II and thiaminase I. Although they perform a similar biochemical reaction, the two groups of proteins lack sequence similarity. Thiaminase II (known as TenA in bacteria) uses H₂O as a nucleophile and is widely distributed in bacteria and archaea (13, 14). Homologs are also found in fungi (15) and plants (16). These intracellular enzymes function in the recycling of base-

degraded thiamin, which contains an intact HMP moiety conjugated to a formylamino group. In *Bacillus halodurans* base-degraded thiamin is imported into the cell by the ThiXYZ transport system and then deformylated by YlmB (17). TenA catalyzes the substitution of the amino group with a hydroxyl from H₂O, effectively restoring HMP (17).

Thiaminase I enzymes are much less widely distributed, occurring only in a small number of phylogenetically unrelated bacteria (2), and two amoeba (18). Thiaminase I gene homologs are coded for in the genomes of a number of human pathogens including *Clostridium botulinum* strains, *Burkholderia pseudomallei*, and *Naegleria fowleri*, and their close, non-pathogenic relatives *C. sporogenes*, *B. thailandensis*, and *N. gruberi*, as well as in three *Paenibacillus* species (2). The bacterial enzymes are secreted through the general secretory pathway (19). Thiaminase I can use a variety of organic nucleophiles including pyridine and cysteine for the base-exchange reaction (20-22). Thiaminase I activity in culture is repressed by high levels of thiamin (23, 24). Crystal structures uncovered the mechanism of the substitution by revealing the structure of the active site which includes a catalytic cysteine (19, 25). Further, they revealed that thiaminase I is structurally similar to and shares a common ancestor with TbpA (ThiB), the thiamin binding protein used for thiamin transport in *E. coli* (26). Biochemical and genomic analysis revealed that the thiaminase I gene of *C. botulinum* A ATCC 19397 is in an operon with the biosynthesis genes for the HMP analog bacimethrin, which when combined with THZ, forms the antivitamin 2'-methoxythiamin (27). Despite being located in this operon, its role if any in the synthesis of bacimethrin is unknown (27). In *Paenibacillus* species, it is found in a conserved operon including thiamin salvage and synthesis genes *thiM*, *thiD*, and *thiE* and possibly genes for bacimethrin synthesis. Despite these biochemical, mechanistic, and genomic analyses, the biological function

of thiaminase I secreted by these bacteria remains a great mystery in our understanding of thiamin metabolism.

Thiaminase I has been linked to thiamin deficiencies and death in animals. Despite only being known to be produced by microorganisms, thiaminase I activity has been found in metazoans and plants (20), however, in many cases, it is unclear if these thiaminase I enzymes are produced endogenously or by their microbiota. Thiaminase I induced deficiencies affect economically important domesticated animals, the classic case being Chastek paralysis in which domesticated foxes and mink were fed raw fish high in thiaminase activity (28-30). Sheep grazing on plants with high thiaminase I activity leads to polioencephalomalacia (31), but this can also be triggered by the overproduction of thiaminase I by *C. sporogenes* in the rumen of sheep and other ruminants (32). Thiaminase I induced thiamin deficiency can affect wild populations of animals as well, including alligators (33) and salmonid fish. Consumption of clupeid fish with high thiaminase I activity by salmonids results in Early Mortality Syndrome in which their offspring are thiamin deficient and die in the early life stages (34, 35). It is unclear if the microbiota is responsible for the high thiaminase I content of clupeid fish, as in some cases *Paenibacillus thiaminolyticus* was isolated from the viscera of these fish (36), and in other cases it was not (37). Thiamin deficiency can also arise in humans that consume raw diets high in thiaminase I activity (38). The first known thiaminase I producing bacteria were isolated from the feces of an individual experiencing thiamin deficiency (20).

Due to its occurrence with genes involved in thiamin salvage and biosynthesis in some genomes, it has been hypothesized that thiaminase I plays a role in thiamin salvage (26). Here we use *B. thailandensis*, a thiaminase I producing soil bacterium found in rice paddies and surface water, to test this hypothesis. Since *B. thailandensis* is capable of synthesizing thiamin, we

generated thiamin auxotrophic mutants incapable of synthesizing either HMP or THZ and examined the impact of thiaminase I production on growth in defined media. We found that thiaminase I, either produced by the cell or added to the growth medium, provided a growth advantage to thiamin auxotrophs. Furthermore we found that thiamin auxotrophs grew better on exogenous HMP and THZ than when they were provided thiamin. Thiaminase I also allowed for growth of thiamin auxotrophs on the phosphorylated forms of thiamin, and allowed growth on certain thiamin analogs. This study is the first to ascribe a function to the elusive thiaminase I enzyme. We propose to name the thiaminase I gene locus (BTH_II1306) *thiA* due to its current inconsistent nomenclature and its apparent role in thiamin salvage.

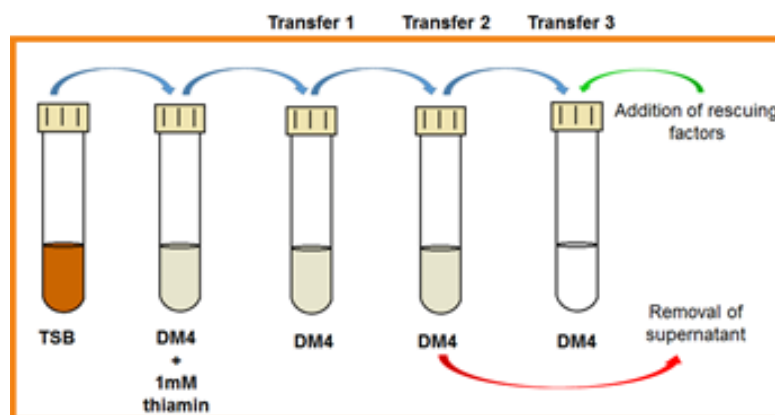


FIG 1 Experimental overview of the serial transfer experiments. All strains were originally grown in TSB, then after 24 hrs, transferred to DM4 with 1 mM thiamin. All cultures were inoculated to a starting OD₆₀₀ of 0.025. Cultures were then serially transferred to fresh DM4 every 24 hrs. In the illustration shading of the liquid medium in tubes indicates growth, while clear tubes indicate no growth. Supernatant was taken from tubes for the rescue experiments, as indicated by red arrow, during the final transfer that showed growth. Addition of rescuing factors, indicated with a green arrow, shows the transfers when treated supernatant, BcmE (thiaminase I), thiamin, precursors, or analogs were added to the medium.

4.4 RESULTS

Thiaminase I confers a growth advantage to thiamin auxotrophs

To test if thiaminase I plays a role in thiamin salvage or biosynthesis, we generated *B. thailandensis* thiamin auxotroph mutants in wildtype strain E264 and BT10432, an isogenic strain

with a transposon inserted into *thiA* (thiaminase I). Both strains are capable of synthesizing thiamin. When assayed for ThiA activity under growth conditions that normally produce substantial thiaminase I activity in strain E264, we found none in BT10432 (not shown). In both backgrounds, *thiC* was disrupted thus blocking the synthesis of HMP, or *thiG* was disrupted thus preventing the formation of THZ, or both genes were disrupted (Table 1). The 8 strains, E264, E264-*thiC*, E264-*thiG*, E264-*thiCthiG*, BT10432, BT10432-*thiC*, BT10432-*thiG*, and BT10432-*thiCthiG*, were serially transferred every 24 hrs from TSB to defined medium (DM4) containing 1 mM thiamin, to thiamin-free DM4 until they would no longer grow. A schematic of experimental design is shown in Fig. 1.

After the first transfer into thiamin-free DM4, a growth defect was observed in the BT10432-*thiC* and BT10432-*thiCthiG* mutants, as both strains had a significantly lower OD₆₀₀ at 24 hrs (Fig. 2A). The only other significant difference observed at this time point was E264-*thiC* had a higher OD₆₀₀ than E264-*thiG*. This suggests that a defect in thiazole synthesis, via ThiG, is tolerated better than a defect in the ability to make HMP. Plus an intact *thiA* provides a growth advantage to thiamin auxotrophs.

In the next serial transfer, none of the BT10432 thiamin auxotrophs were able to grow whereas all the E264 thiamin auxotrophs were able to grow (Fig 2B). The E264-*thiC* and E264-*thiCthiG* strains had longer lag in the media in comparison to the other strains that were able to grow. This demonstrated that ThiA was providing the E264 strains with a growth advantage compared to isogenic auxotrophs without thiaminase I. Consistent with what occurred in the BT10432 mutants, in the next serial transfer, the E264-*thiC* and E264-*thiCthiG* mutants were unable to grow, while the E264-*thiG* mutant was able to grow (Fig. S1A), recapitulating that *B. thailandensis* thiamin auxotrophs are more sensitive to the loss of the ability to make HMP, than

THZ. The E264-*thiG* mutant was unable to grow in the next serial transfer, while the strains capable of thiamin synthesis were able to grow (Fig S2A). To ensure that thiaminase I was produced by the E264 thiamin auxotrophs during the transfers, supernatant was assayed for thiaminase I activity. Peak thiaminase I activity was detected at 18 hrs in transfer 1 in DM4, while peak activity was detected at 12 hrs in transfers 2 and 3. Taken together, these results demonstrate that thiaminase I provided thiamin auxotrophic *B. thailandensis* strains a growth advantage, and these mutants are more sensitive to the loss of the ability to synthesize HMP rather than THZ. Further, the differences in growth phenotypes of the different types of mutants rules out the possibility that ThiA is synthesizing thiamin in the reverse reaction, as growth would be consistent between thiamin auxotroph strains.

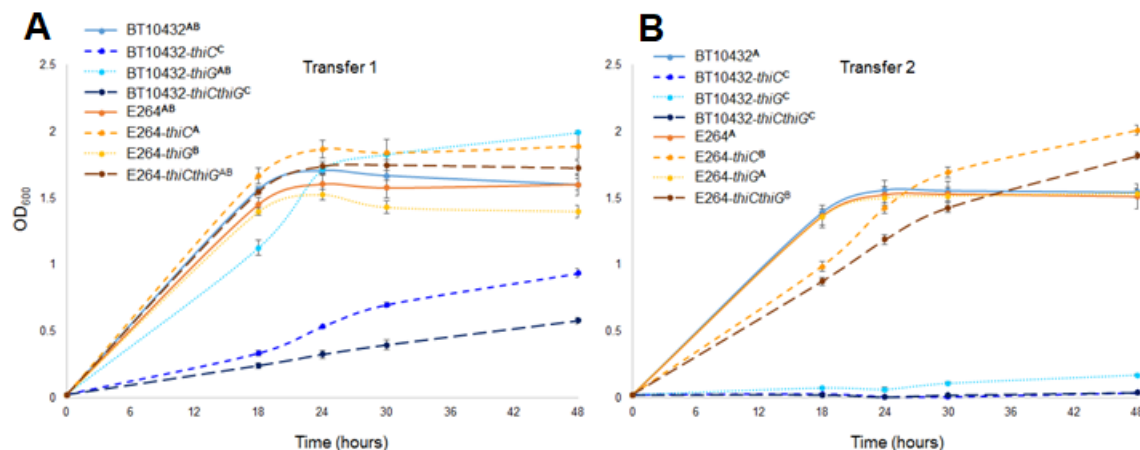


FIG 2 Survival of *B. thailandensis* strains in thiamin-free DM4. These data illustrate the decline of thiamin auxotrophs that cannot make thiaminase I. (A) Indicates the growth of the 8 strains in the first transfer into thiamin-free DM4. Strains incapable of making ThiA and defective in synthesizing HMP or HMP and THZ are the first to show a significant reduction in growth ($p < 0.0001$). (B) After the second transfer in thiamin-free medium, the strain defective in both ThiA and THZ production drops out. The growth curves are displayed as the average of three replicate cultures with standard error bars at each time point. Strains with the same superscript designations next to the names in the legend indicate no significant difference (for example AB is not significantly different from A or B). For all of the pairwise comparisons described here $p < 0.0001$.

Thiaminase I is the rescuing factor allowing for growth

Our first set of experiments indicated that thiaminase I was somehow allowing for growth of thiamin auxotrophs in DM4 with thiamin diluted out, but to better understand this, we examined if we could rescue growth of the BT10432-*thiC* auxotroph with ThiA containing conditioned supernatant. To do this, we collected supernatant from BT10432-*thiC* and E264-*thiC* cultures at the penultimate transfer before each strain could no longer grow (see Fig 1. for experimental setup) and filter-sterilized it. The cell-free supernatant was combined with fresh thiamin-free DM4 in a 1:1 mixture. We then used this as the medium for the terminal transfer of BT10432-*thiC* and transfer 2 into DM4 of E264-*thiC*. In both cases, the uninoculated 1:1 mixture exhibited no growth (data not shown). The BT10432-*thiC* treated medium did not rescue growth of BT10432-*thiC*. When added to E264-*thiC* it did not inhibit growth in fact, this culture had a significantly higher OD₆₀₀ at 24 hr in this mixture possibly due to a faster recovery from the lag phase as the cultures may have had a higher thiamin content due to the reduced growth of the BT10432-*thiC* mutant when the sample was collected (Fig. 3A). The lack of growth of the BT10432-*thiC* strain indicated that cell lysis releasing thiamin or precursors into the culture supernatant during growth was not supporting growth of the auxotrophs. When BT10432-*thiC* was grown in the E264-*thiC* treated mixture, growth was rescued, though the OD₆₀₀ was significantly lower than the E264-*thiC* strains (Fig 3A). This finding indicated that the presence of ThiA, or its transformation of the medium components, was enough to rescue the growth of BT10432-*thiC* auxotroph.

To unambiguously demonstrate that thiaminase I was the rescuing factor, we purified BcmE, the thiaminase I from *C. botulinum* (25) and added it to DM4 at the terminal transfer step of the serial transfers for all auxotrophic strains. BcmE rescued growth for all of the BT10432 strains (Fig 3B). This suggests that thiaminase I is able to provide the auxotrophic mutants with

their respective precursors to generate thiamin. Further, it demonstrates that the thiaminase I products do not provide a species specific benefit. When BcmE was added to the E264 auxotrophs, it did not rescue growth in the *thiC*⁻ and double mutant cultures (Fig 3C). Although there was a bit of growth of E264-*thiG*, this was not at the same level as the BT10432-*thiG* strain or the other *thiA*⁻ strains (Fig 3B). This is likely due to the presence of *B. thailandensis* ThiA already acting on the media and transforming it in the previous transfers. At this stage, any source of thiamin in the media were likely approaching exhaustion so there was not substrate for the exogenously added BcmE to act upon.

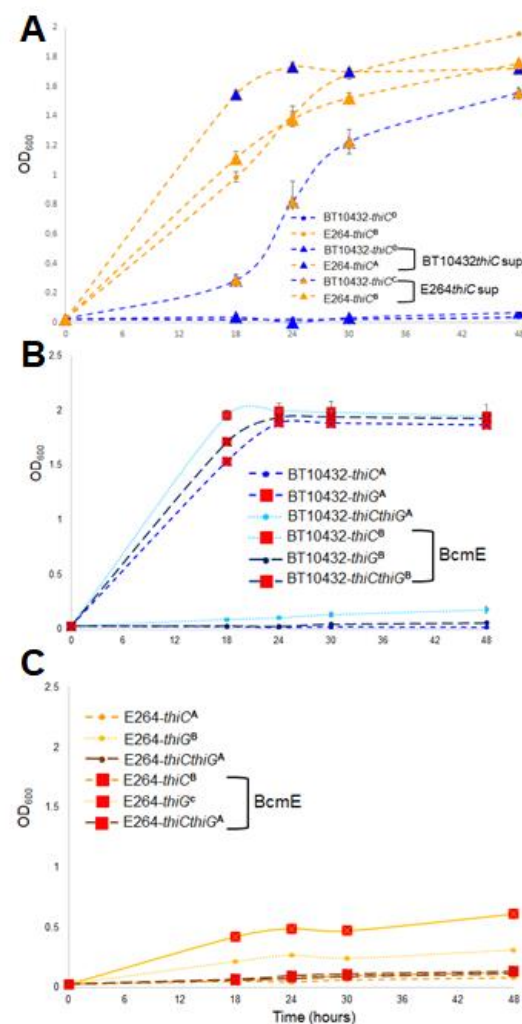


FIG3 Thiaminase I rescues of thiamin auxotrophs. (A) depicts the growth of BT10432-*thiC* and E264-*thiC* in DM4 supplemented with either BT10432-*thiC*-DM4 supernatant or E264-*thiC*-DM4 supernatant. The ThiA containing supernatant rescued growth, while supernatant lacking ThiA did not rescue growth of BT10432-*thiC* (B) demonstrates that BcmE rescues the growth of the BT10432 thiamin auxotrophs. (C) reveals that the BcmE addition does not substantially rescue growth of the E264 thiamin auxotrophs as there was small growth in the BcmE rescued E264-*thiG* cultures. The growth curves displayed are the average of three replicate cultures with standard error bars at each time point. In each case, growth at 24 hr was compared and significant differences are designated by different superscripts next to the names in the legend. $p < 0.0001$ for all pairwise comparisons in these experiments

Thiaminase I- thiamin auxotrophs preferentially grow on thiamin precursors

Our findings purported that thiaminase I rescued growth by providing the auxotrophs with their necessary precursors to synthesize thiamin. We set out to confirm this assertion by adding back the precursors or intact thiamin to each auxotroph culture. HMP was added for *thiC* mutants, THZ for *thiG* mutants, and both for *thiCthiG* mutants at the terminal transfer step (Fig 4, S2). Strikingly, the BT10432-*thiC* (Fig. 4A) and BT10432-*thiCthiG* (Fig. 4C) cultures reached significantly higher OD₆₀₀ at 24 hr when rescued with the precursors than when the same amount of thiamin was added. In both cases there was a longer delay in growth when thiamin was added back. This lag could be a consequence of preferential uptake of precursors over intact thiamin. In contrast, the E264-*thiC* (Fig. 4B) had a significantly higher OD₆₀₀ at 24 hrs while growing on thiamin rather than HMP. There was no statistically significant difference between growth of the E264-*thiCthiG* mutant on thiamin or both precursors after 24 hr (Fig. 4D). Although growth on thiamin exhibited a greater lag, this culture yielded a higher OD₆₀₀ in stationary phase. This proposes that *B. thailandensis* preferentially uses the substrates of thiamin to grow, rather than thiamin itself. The superior growth (E264-*thiC*) or lack of a difference of the growth (double mutant) when ThiA is present suggests that it functions to degrade extracellular thiamin to its precursors to supplement growth. This appears to be the case for when HMP cannot be synthesized,

however, the BT10432-*thiG* mutant exhibits significantly higher OD₆₀₀ when grown on thiamin rather than THZ (Fig S2A). This could have potentially been a consequence of THZ degradation in the media. Although inconsistent with the HMP and HMP + THZ mutant rescues, it is consistent with our previous findings that the *thiG* mutants are less sensitive to their specific auxotrophy. Further, a thiochrome assay was performed to ensure that the thiamin was not simply degrading in the culture medium and thus making precursors available. Based on the assay, after 24 hrs of incubation, there was no apparent change in thiamin concentration in uninoculated medium, ruling out this possibility. This does not rule out the possibility that thiamin may have degraded in the stock solutions and there may have been a THZ supply from that. Consistent with the other E264 mutant strains, there was no difference in the growth of the E264-*thiG* mutant on THZ or thiamin.

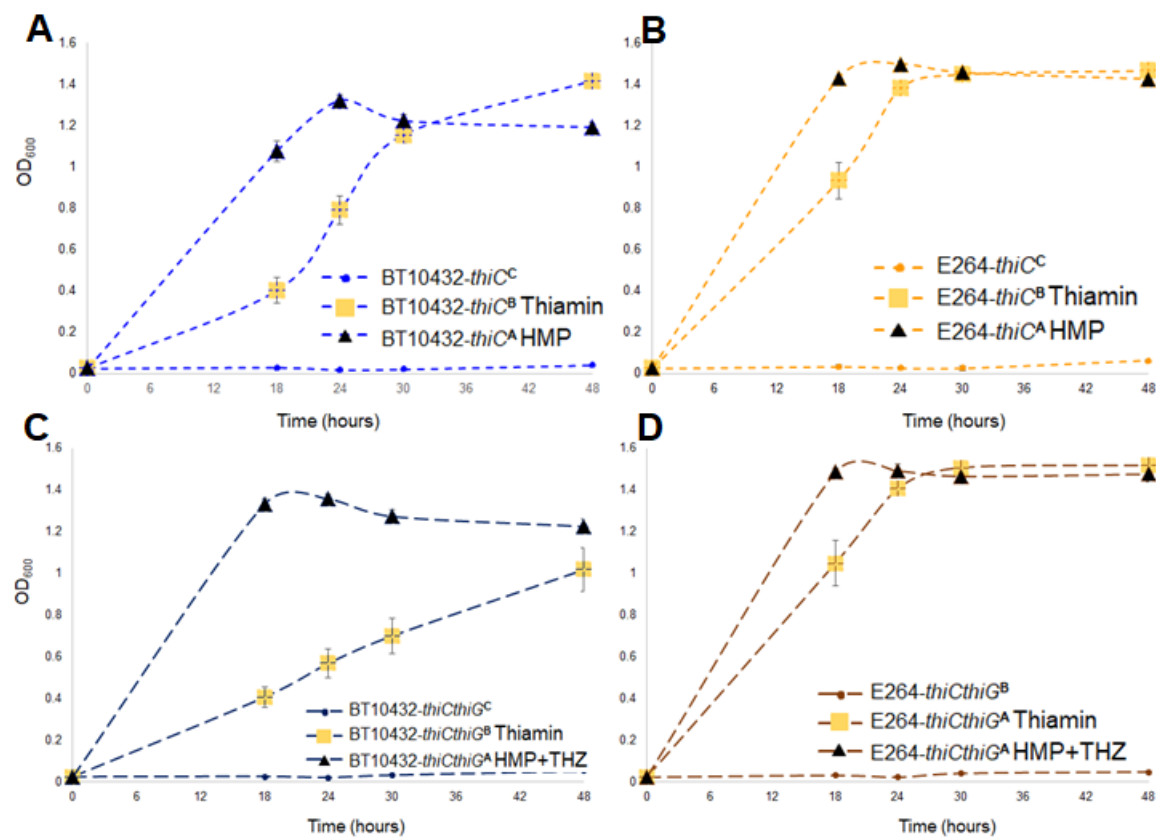


FIG 4 Growth of *B. thailandensis* strains on thiamin and its precursors. In the absence of thiaminase I, thiamin auxotrophs grew better on thiamin precursors than thiamin. (A) The growth curve showing BT10432-*thiC* reaches a significantly higher OD₆₀₀ ($p = 0.0001$) with HMP in

contrast to thiamin. (B) demonstrates E264-*thiC* reaches a significantly higher OD₆₀₀ ($p < 0.0001$) when rescued with thiamin in contrast to HMP. (C) shows BT10432-*thiCthiG* reached a significantly higher OD₆₀₀ ($p = 0.0004$) when rescued with HMP+THZ in contrast to thiamin. (D) shows there was no significant difference between growth on thiamin and HMP+THZ for E264-*thiCthiG* at 24hrs. The growth curves displayed are the average of six replicate cultures for precursors and nine for thiamin additions with standard error bars at each time point. In each case, growth at 24 hr was compared and significant differences are designated by different superscripts next to the names in the legend.

Phosphorylated thiamin compounds rescue growth only when thiaminase I is present

In bacteria, compounds that need to be phosphorylated to activate them are imported in their dephosphorylated form then phosphorylated in the cytoplasm (26). TbpA and ThiT can both bind phosphorylated forms of thiamin, however, TPP is bound with a lower affinity (26) It has also been observed that thiaminase I from *P. thiaminolyticus* is as active against the phosphorylated forms of thiamin as it is thiamin (21, 49). Based on these observations, we hypothesized that thiaminase I may function to recycle the constituents of phosphorylated forms of thiamin, most likely the non-phosphorylated HMP. Supporting our hypothesis, there was no substantial growth when either TMP or TDP was added to the BT10432-*thiC* or BT10432-*thiCthiG* mutants (Fig 5A). However, when either compound was added at the terminal transfer, both E264-*thiC* and E264-*thiCthiG* were rescued (Fig 5B). There was no difference between growth on either compound or between the different thiamin auxotrophs. This suggests that *B. thailandensis* has a mechanism to either dephosphorylate the phosphorylated THZs freed by ThiA or it can import these compounds. Regardless, the THZ-PP would need to either be dephosphorylated to THZ-P for recycling, as THZ-P is combined with HMP-PP by ThiE, or dephosphorylated all the way to THZ and then phosphorylated to THZ-P by ThiM. Consistent with other findings throughout this study, the BT10432-*thiG* mutants did not behave like the other BT10432 mutants, as it was able to grow slowly with TMP, but did not grow substantially with TPP (Fig S3A). This might be a result of the single phosphate of TMP being more labile in media than the pyrophosphate of TPP. The E264-

thiG cells were able to recycle the THZ-P and THZ-PP from TMP and TPP respectively. There was no difference in growth on these compounds (Fig S3B).

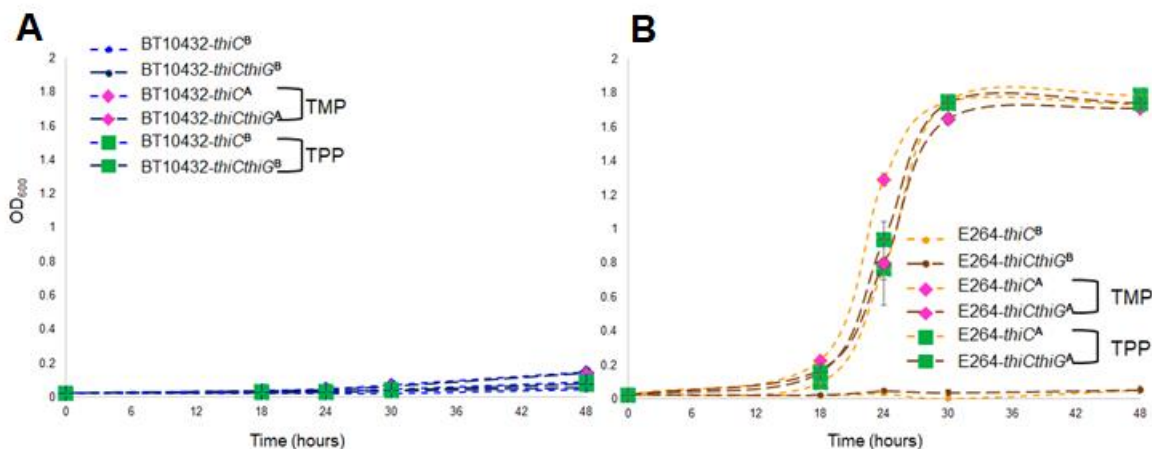


FIG 5 Growth of *B. thailandensis* strains on TMP and TPP. Thiamin auxotrophs can grow when media is supplemented with phosphorylated forms of thiamin but only when the strains produce ThiA. (A) shows there was no substantial growth of BT10432-*thiC* and BT10432-*thiCthiG* on TMP and TPP. When TMP was present there was a significantly higher OD₆₀₀ at 24 hrs than when TPP was added. (B) shows there was substantial growth of E264-*thiC* and E264-*thiCthiG* on TMP and TPP. When TMP is present there was a significantly higher OD₆₀₀ at 24 hrs than when TPP was added. The growth curves displayed are the average of six replicate cultures with standard error bars at each time point. In each case, growth at 24 hr was compared and significant differences are designated by different superscripts next to the names in the legend.

Thiaminase I provides an advantage for growing on certain thiamin analogs

The *P. thiaminolyticus* thiaminase I was shown to be active on certain thiamin analogs such as pyrithiamin (Table S2) (21, 50, 51). This lead us to hypothesize that thiaminase I may function to release portions of these thiamin analogs that could then be imported and used to synthesize thiamin intracellularly. To test this, we added different thiamin analogs to DM4 at the terminal transfer step. Analogs examined and their structures are shown in Table S2. These include compounds such as pyrithiamin and oxythiamin, which contain an intact HMP or THZ respectively, as well as thiamin disulfide and dibenzoyl thiamin, which have an intact HMP with a bulky side chain that contains a ring-open form of THZ. We found that when pyrithiamin was

added to the medium E264-*thiC* had a significantly higher OD₆₀₀ at 24 hrs than the BT10432-*thiC* (Fig 6A), however, by 48 hours, the BT10432 *thiC* mutant exceeded the growth of E264-*thiC*. This suggests that early in the growth of E264-*thiC*, thiaminase I is able to free HMP through pyrithiamin degradation giving it an initial boost in growth, mitigating the lag phase experienced in the BT10432-*thiC* culture. Then it appears that either abiotic or biotic degradation through another mechanism provides the BT10432 mutant with HMP. In addition, the BT10432-*thiC* mutants were unable to grow substantially using the thiamin disulfide and dibenzoyl thiamin analogs at low levels or rates, however, after a long lag the E264-*thiC* mutants were able to grow on both analogs (Fig 6A). This suggests that thiaminase I might be able to degrade thiamin disulfide and dibenzoyl thiamin, releasing HMP, which can then be used by the cell. However, previous reports have described thiamin disulfide as recalcitrant to thiaminase I (51) and we were unable to detect any thiaminase I activity using BcmE with these compounds. It could be that the kinetics of the reaction are very slow and the growth lag was scored as no-growth in other bacteria. A more likely interpretation is that thiamin disulfide was abiotically converted to thiamin (52), and thiaminase I then acted on the thiamin. The slow kinetics of abiotic degradation could account for the initial lag in growth. For dibenzoyl thiamin, the sulfur was likely reduced abiotically in these cultures, but this reduced form would still have a bulky benzoyl group attached to the thiamin. Despite this, thiaminase I appears to be acting on this compound, as only E264-*thiC* showed considerable growth.

Next we added pyrithiamin and/or oxythiamin to DM4 at the terminal transfer in the BT10432 and E264 double mutants (Fig 6B). Although little growth was observed at 24 hrs, the E264-*thiCthiG* mutant had significantly more growth when pyrithiamin and oxythiamin were added in concert. By 30 hrs, the OD₆₀₀ nearly doubled and the culture reached a high density by

48 hrs. In the BT10432 background, the OD₆₀₀ only reached ~0.5 at 48 hrs. This demonstrated that ThiA was likely cleaving both analogs, freeing the HMP from pyrithiamin, and the THZ from oxythiamin. These were then salvaged by *B. thailandensis*. The fact that there is some growth of the BT10432-*thiCthiG* mutant suggests that there may be abiotic degradation of the compounds, or *B. thailandensis* has a less efficient way of converting these analogs into usable forms as opposed to ThiA. This was further obfuscated by the *thiG* mutants (Fig S4). In both the E264 and BT10432 backgrounds, there was no significant difference in growth when oxythiamin was added, although the strain without ThiA showed a growth lag at 18 hr. This suggests that *B. thailandensis* has another mechanism for detoxifying this compound and restoring the THZ. In *E. coli*, L-aspartate or L-glutamate is used to generate a thiaminosuccinate intermediate for oxythiamin conversion to thiamin (53) but *B. thailandensis* does not appear to be using this mechanism as oxythiamin alone does not rescue the the double mutants. This called into question if ThiA was acting on oxythiamin in the E264-*thiCthiG* oxythiamin and pyrithiamin rescue or if THZ is acquired in a different fashion. Our thiaminase I assay did not demonstrate activity on this compound supporting the idea that *B. thailandensis* has another mechanism for detoxifying oxythiamin. ThiA provides a growth benefit to the *thiG* mutants when thiamin disulfide or dibenzoyl thiamin was added, as they reached a significantly higher OD₆₀₀ when compared to the BT10432-*thiG* mutants (Fig. S4) and its presence reduced the lag time for growth on these compounds. These compounds have a ring-opened THZ, and the ability for these compounds to rescue the growth of the E264-*thiG* mutant is either a consequence of the abiotic conversion to thiamin, the abiotic circularization of freed THZ, or an alternative mechanism for regeneration of the THZ ring.

FIG 7 Thiamin metabolism in *B. thailandensis* E264. Wild type *B. thailandensis* is capable of synthesizing thiamin, depicted in this illustration of a *B. thailandensis* E264 cell. The thick black line around the cell represents the cell envelope. Proteins involved in the synthesis of thiazole are represented by ovals in different shades of green and arrows are in dark green. Proteins involved in the synthesis of the pyrimidine moiety are represented by triangles in different shades of blue, arrows are in dark blue, and the CytX transporter for HMP uptake is labeled. The proteins that combine the moieties to make TMP and TPP are depicted as rounded squares in shades of red. Extracellular thiaminase I, ThiA, is depicted as orange octagons acting on both thiamin and pyrithiamin. The X on the thiamin molecule represents the side chain of either thiamin, TMP, or TPP. The R on the HMP molecule represents the diversity of organic nucleophiles attached to it. Unclear pathways are represented as dashed lines. In our model, thiamin disulfide is abiotically converted to thiamin, which ThiA is acting on, however, it is possible that ThiA is acting on this molecule as well, at a slow rate depicted by the dashed orange line. In the model dibenzoyl thiamin loses the benzoyl group attached at the sulfur abiotically, and the compound is then acted on by ThiA. It is unclear how the second benzoyl group is removed. It is unclear if HMP-R is processed extracellularly and imported as HMP, or if it is imported by an unknown transporter (white oval transporter in HMP section labeled with ?) and then processed by an unknown enzyme (white triangle labeled with ?) to HMP. In the thiazole section, it is unclear if it is processed extracellularly to remove phosphates or intracellularly, and if it becomes carboxylated after import.

4.5 DISCUSSION

In this study, we used the genetic model *Burkholderia thailandensis* to demonstrate that the enigmatic enzyme thiaminase I plays a role in thiamin salvage. We found that thiamin auxotrophs with ThiA are able to survive more serial dilutions in defined medium than the same auxotrophs that lack the ability to make thiaminase. This survival advantage is likely due to the ability of ThiA to generate thiamin precursors which are preferentially used over thiamin by *B. thailandensis* thiamin auxotrophs. Further, thiaminase I allows *B. thailandensis* to salvage phosphorylated forms of thiamin, as well as some thiamin analogs. Our proposed model for the role thiaminase I plays in thiamin biosynthesis and salvage is provided in Figure 7, although there are some details that still need to be resolved. First, it is unclear if the HMP with a covalently attached nucleophile, HMP-R, can funnel directly in to the thiamin biosynthesis pathway, or if it has to be processed first to remove the R group. If it is processed, it is unknown if this occurs extracellularly or intracellularly. It is also unclear if HMP-R can be imported through the CytX

HMP transporter, or if there is a transporter specific for it. In the case of THZ, it is unknown how THZ enters the cell in *B. thailandensis*, as no *thiW* homolog has been identified. Further, the THZ released by ThiA from thiamin is decarboxylated in comparison to the THZ carboxylate tautomer precursor. For its recycling, this thiaminase I product either needs to be carboxylated, or it might be acceptable as is. The lack of the charged carboxyl group may aid in its diffusion across the membrane, and this could partially account for why the E264-*thiG* mutants are able to survive longer than the E264-*thiC* mutants. When recycling TMP and TPP, it is unclear if the freed phosphorylated or pyrophosphorylated THZ molecules need to be dephosphorylated before entering the cell. In terms of recycling the thiamin analogs, the intact moieties can be recycled, while the damaged or different moieties can be avoided. In the case of the thiamin disulfide and dibenzoyl thiamin analogs, it is unclear if ThiA acted on the compounds first or if they abiotically converted to thiamin, which was then processed by ThiA.

One confounding finding in our results was the growth of BT10432-*thiC* on pyrithiamin. Growth on this compound suggests multiple possibilities; i) *B. thailandensis* is able to use pyrithiamin as a cofactor, which is highly unlikely since the *thiCthiG* mutants were not rescued by pyrithiamin. ii) Pyrithiamin was slowly degraded in the medium, freeing HMP for import and salvage. iii) The most likely explanation is that the pyrithiamin was contaminated with low levels of HMP. Carini and colleagues detected HMP contamination in their thiamin and 4-amino-5-aminomethyl-2-methylpyrimidine stocks (8), and we might have encountered a similar problem with our stocks. ThiA still provided an initial advantage by acting on this compound, but some levels of HMP contamination may have been enough to eventually allow for the growth of the BT10432-*thiC* mutant. *B. thailandensis* E264 is a very capable thiamin producer and does not appear to encode for any known thiamin transporter, making it unclear how thiamin would enter

the cell. The growth exhibited by the BT10432 strains when rescued with thiamin (Fig. 4), might also be due to precursor contamination, as could the growth of the *thiG* mutants on oxythiamin. It seems unlikely that *B. thailandensis* E264 has a unique strategy for detoxifying oxythiamin directly to thiamin. If this was the case, we should have observed growth of the *thiCthiG* mutants when supplied oxythiamin alone (Fig. 6). Rather this demonstrated that THZ is either coming from oxythiamin through oxythiamin degradation (abiotic or an unknown biotic mechanism) or there is likely some THZ contamination. THZ contamination may also be occurring at different levels and might account for the differences in growth between the HMP and THZ auxotrophs.

The biological significance of the preference of *B. thailandensis* to grow on thiamin precursors rather than thiamin itself is unclear. However, this phenomenon has been observed in other bacteria. Members of the SAR11 clade are thiamin auxotrophs as they lack *thiC*, and in these bacteria, thiamin-limited growth is not rescued by the addition of thiamin to media, rather it only occurs when HMP is present (8). This might be an evolutionary consequence of their symbiosis with cyanobacteria which exude HMP, and the fact that HMP concentrations are higher than thiamin near the deep chlorophyll maximum zone (8). Unlike the SAR11 group, *B. thailandensis* is a soil bacterium originally isolated from rice field soil and water, where the concentrations of thiamin, HMP, and THZ are not known. Thiamin adsorbs to clay particles in soil, and the strength of adsorption is pH dependent as thiamin can have 2 positive charges, one on the HMP moiety at position 1 and one in the amine group of the THZ moiety (44, 45). This adsorption may reduce the bioavailability of thiamin, forcing soil microbes to evolve unique strategies for acquiring environmental thiamin. It is possible that by degrading thiamin to its precursors, *B. thailandensis* E264 reduces the strength at which thiamin can adsorb to clay and other constituents of soil, making them more biologically available for uptake. These can then be salvaged for synthesis.

Aside from adsorbing to clay particles, thiamin is also labile in soil. Factors such as temperature, pH, the presence of bisulfites and other inorganic bases, metal complexes, and ultraviolet radiation all can contribute to the abiotic destruction of thiamin (46). Jenkins and colleagues demonstrated that thiamin degradation occurs not only in alkaline, but also neutral soils (17), these factors may influence the ecology of *B. thailandensis* E264 thiamin metabolism in multiple ways. First, thiaminase I may be active on multiple naturally occurring thiamin breakdown products present in its environment. This would free intact precursors, most likely the HMP as THZ is less stable, for the purposes of recycling, essentially serving an analogous function to TenA. There are differences in the specificities between the 2 enzymes (41), making it more likely that ThiA can accommodate a greater diversity of thiamin breakdown compounds, especially since the THZ moiety, which TenA has a specificity for binding, is more prone to degradation. The ability of ThiA producers to recycle these compounds was of particular concern due to the addition of the organic base to HMP in the base-exchange reaction. Our results clearly demonstrate that the addition of the organic base to HMP does not prevent its ability to be recycled.

The second influence on *B. thailandensis* thiamin ecology is that processing thiamin and thiamin analogs extracellularly might serve as a means for preempting the cell against toxic analogs, as it would prevent the import of these intact compounds, and their subsequent phosphorylation and competition with TPP for binding of TPP-dependent enzymes. Organisms take preemptive measures to prevent the utilization of multiple groups of toxic metabolites and cofactors (47). Recently, a Nudix family hydrolase was shown to be found in thiamin biosynthesis operons in bacteria, plants, and fungi that preempts these organisms from incorporating toxic analogs (48). This enzyme removes the phosphate groups from oxythiamin and oxothiamin, preventing their incorporation into TPP-dependent proteins (48). In *B. thailandensis*, there is a

Nudix family protein present in the ThiA operon, and may serve this function. ThiA could be another member of this preemption process and it provides the cell with the added bonus of being able to recycle usable thiamin moieties from these damaged compounds. Our understanding of thiamin metabolism, particularly thiamin salvage and detoxification, is still limited, as there are new proteins that appear seemingly essential to this metabolism, but have functions not yet elucidated (49). Although our study supports a role for thiaminase I in thiamin salvage, more investigations need to be made to further clarify its roles.

Converting thiamin to its precursors and using them instead of thiamin may provide a competitive advantage to *B. thailandensis* in a non-repair manner. ThiA activity reduces the already scarce amount of thiamin available in the environment for competing bacteria to use, but would increase the amount of precursors present. This may give *B. thailandensis* a leg up when competing against thiamin auxotrophs that can only import thiamin. It would also potentially provide a boost when competing against other thiamin-capable organisms that are unable to salvage thiamin from its precursors. By salvaging thiamin from precursors in the environment *B. thailandensis* can bypass multiple steps in the thiamin synthesis pathway, thus reducing the energy input to make thiamin. By depleting thiamin, its competitors would have to synthesize thiamin through the full pathway putting them at a growth disadvantage. If these competitors are normally able to import thiamin from the environment, they would no longer be able to, and thus be at a disadvantage. The ability to recover HMP from phosphorylated thiamin would give *B. thailandensis* an advantage against those bacteria that lack a TbpA, as this provides it with a source of HMP that would be generally unusable by other bacteria due to their inability to import phosphorylated thiamin. *B. thailandensis* E264 has an arsenal of different competition factors to compete against other bacteria (50), and this may be another component of its repertoire.

Although not known to be pathogenic to humans, *B. thailandensis* is able to infect mammals at high doses (50-52) and insect models (53, 54). Thus thiaminase I may be another competition/pathogenicity factor for interacting with animals. Since animals are unable to synthesize thiamin, the production of ThiA would convert any free thiamin into a form *B. thailandensis* can utilize but its host cannot during infection. Further, HMP has been characterized as toxic to rats as it interferes with enzymes that require pyridoxal 5'-phosphate (55). Therefore production of HMP may have a three-fold effect; i) it would supplement *B. thailandensis* growth, ii) reduce thiamin levels in the host, and iii) generate a compound potentially toxic to the host. More studies are needed to substantiate these hypotheses.

Our results shed light on a long-standing mystery in thiamin metabolism. The role of thiaminase I has remained enigmatic, and here we provide the first evidence supporting a role in thiamin salvage. Future studies are needed for a comprehensive understanding of thiamin metabolism of *B. thailandensis*. Our study also raises caution with using thiaminase I as a tool to eliminate thiamin from media (1). Other organisms may be like *B. thailandensis*, where they prefer to grow on the thiamin precursors generated by the enzyme. This could possibly confound results, consequently whenever possible, using a defined medium instead would give a more interpretable answer if a microorganism can grow in a thiamin-free environment.

4.6 MATERIALS AND METHODS

Bacteria and yeast strains and growth conditions. Strains are summarized in Table 1. *B. thailandensis* strains were maintained in tryptic soy broth (TSB) and grown at 27°C, while *E. coli* strains were grown in LB at 37°C. *Saccharomyces cerevisiae* was grown YPD media (56) at 27°C. Chemically defined medium (DM4) was generated based on M9 medium (57), however, it was

buffered by 0.1 M MES, pH 6.0 and contains 10 mM FeSO₄, 9.5 mM NH₄Cl, 0.276 mM K₂SO₄, 0.5 μM CaCl₂, 0.525 mM MgCl₂, 50 mM NaCl, 1.32 mM K₂HPO₄, 1% (v/v) Vitamin Supplement (ATCC MDVS) excluding the thiamin, 1% (v/v) Trace Mineral Solution (ATCC MD-TMS), 0.1% (w/v) glucose, and 12.5% (v/v) of all 20 amino acid mix based on published concentrations (58). Since thiamin adsorbs to glassware (59), for generation of DM4, we took precautions to ensure there was no thiamin contamination. When autoclaving was unnecessary, plasticware was used for storage of media components, but when glassware was necessary, it was washed with Fisherbrand Cleaning Solution (Fisher Scientific SC88-500), rinsed 10X with tap, then 10X with deionized H₂O, and baked overnight at ~200°C. Since thiamin is base-labile, a 0.1 M NaOH wash was conducted, followed by the same rinsing and baking procedure.

Table 1 Strains used in this study		
Strains or plasmid	Genotype	Source
Organism strain		
<i>E. coli</i>		
DH5α		Invitrogen
BLD21	overexpression strain	Invitrogen
<i>B. thailandensis</i>		
E264	wild-type	(60)
BT10432	E264 <i>thiA</i> ::ISlacZ-PrhaBo-Tp/FRT	(61)
BT10432- <i>thiC</i>	BT10432 $\Delta thiC$:: <i>tetR</i>	This study
BT10432- <i>thiG</i>	BT10432 $\Delta thiG$:: <i>cat</i>	This study
BT10432- <i>thiCthiG</i>	BT10432 $\Delta thiC$:: <i>tetR</i> $\Delta thiG$:: <i>cat</i>	This study
E264- <i>thiC</i>	E264 $\Delta thiC$:: <i>tetR</i>	This study
E264- <i>thiG</i>	E264 $\Delta thiG$:: <i>cat</i>	This study
E264- <i>thiCthiG</i>	E264 $\Delta thiC$:: <i>tetR</i> $\Delta thiG$:: <i>cat</i>	This study
<i>S. cerevisiae</i>		
InvSc1	<i>MATa/MATα leu2/leu2 trp1-289/trp1-289 ura3-52/ura3-52 his3-Δ1/his3-Δ1</i>	(62)
Plasmids		
pMQ87	<i>CEN6, ARSH4, aaC1, lacZα, oriT, oriColE1, gent^r</i>	(62)
pMQ87 <i>thiC</i>	pMQ87 $\Delta thiC$:: <i>tetR</i>	This study
pMQ87 <i>thiG</i>	pMQ87 $\Delta thiG$:: <i>cat</i>	This study
pET-28 <i>bcmE</i>	over expression vector containing <i>bcmE</i>	(25)

Mutant construction. To generate the *thiC*⁻ and *thiG*⁻ mutants we generated plasmids containing *thiC* (BTH_I2844) and *thiG* (BTH_I3006) deletion-disruptions, using the *S. cerevisiae* recombineering methods generated by Shanks and colleagues (56, 62). Briefly, the 5' end of the *thiC* gene was amplified from E264 genomic DNA, using the pMQ87-*thiC* 5' primer set and the 3' end of *thiC* was amplified using the pMQ87-*thiC* 3' primer set (Table S1). The *thiC* gene was disrupted with a tetracycline resistance gene amplified using the tetracycline primer set (Table S1). These fragments were purified using the QIAquick PCR Purification Kit (QIAGEN) and added along with linearized pMQ87 in the transformation solution (56) containing the yeast. Yeast was grown on the selective uracil drop out media (63). Plasmids were extracted from transformant colonies and DH5α *E. coli* was transformed following the chemical transformation protocol of (64), and plated on 10 μg/mL gentamicin containing LB plates. Transformants were screened using colony PCR to ensure that the construct was the expected size, and if so the plasmid were extracted. Generation of the *thiG* disruption was created the same way, except *thiG* specific primers were used (Table S1) and *thiG* was disrupted with the chloramphenicol resistance gene (*cat*).

Generation of *B. thailandensis* mutants. To generate the *B. thailandensis* disruption mutants, we followed the protocol generated by Thongdee and colleagues, which takes advantage of *B. thailandensis*' natural competence (65). Competent E264 and BT10432 cells were generate as described (65). The *thiC* disruption fragment was PCR amplified from the pMQ87-*thiC-tetR* construct using the *thiC*5'F and *thiC*3' R primer pair, cleaned and added to the competent *thiA*⁺ and *thiA*⁻ strains to generate E264-*thiC* and BT10432-*thiC*. The E264 transformation culture was plated on LB with 50 μg/mL tetracycline and the BT10432 transformation cultures were plated on LB with 50 μg/mL tetracycline and 100 μg/mL trimethoprim. Recombination of the knockout

constructs into the chromosome was verified via PCR with confirmation primer sets (Table S1). The *thiG* disruption mutants (E264-*thiG* and BT10432-*thiG*) were generated the same way except the *thiG* specific primer pair (Table S1) was used to generate the PCR fragment for transformation. The E264 transformation culture was plated on LB with 256 µg/mL chloramphenicol and the BT10432 transformation cultures were plated on LB with 256 µg/mL chloramphenicol and 100 µg/mL trimethoprim. To generate the E264-*thiCthiG* double disruption mutant, genomic DNA was isolated from E264-*thiC* and competent E264-*thiG* was transformed with it, and plated on LB containing 256 µg/mL chloramphenicol and 50 µg/mL tetracycline. Colonies were screened the same way as described previously. To generate the double mutant in the *thiA*⁻ background, E264-*thiCthiG* was transformed with BT10432 genomic DNA, and colonies were plated on LB with 256 µg/mL chloramphenicol, 50 µg/mL tetracycline and 100 µg/mL trimethoprim.

Serial transfer experiments. For each experiment *B. thailandensis* strains were plated from frozen stocks on TSB at 30°C. After 3 days of growth, individual colonies for each strain were grown in TSB tubes individually overnight at 27°C with shaking at 200 rpm. The OD₆₀₀ was read with a Pharmacia Biotech Ultrospec 2000 spectrophotometer (SCINTEK Instruments) and subcultured in 3 mL of DM4 supplemented with 1mM thiamin-HCl at a starting OD₆₀₀ of 0.025. Sterile, plastic 13 mL tubes were used for all cultures containing DM4 and each culture was grown at 27°C with shaking at 200 rpm. Growth was recorded spectrophotometrically 18, 24, 30, and 48 hours. At 24 hours, the cultures were subcultured into DM4 without any thiamin at the starting OD₆₀₀ of 0.025, and growth was recorded the same way. The serial dilutions were repeated until the auxotrophs could not grow. E264 and BT10432 were always grown as positive controls for each experiment, depicted in Figure 1. For the original serial transfer experiments, all 8 strains

were used; E264, E264-*thiC*, E264-*thiG*, E264-*thiCthiG*, BT10432, BT10432-*thiC*, BT10432-*thiG*, and BT10432-*thiCthiG*, n=6. OD₆₀₀ at 24 hrs for each experimental grouping was statistically compared in JMP Pro 12.0.1 via an analysis of variance (ANOVA). If there were significant differences, pairwise comparisons were made using a Student's *t*-test. P-values were Bonferroni corrected based on the number of comparisons made.

Thiaminase I activity. To assay thiaminase I activity in cultures grown in DM4, 1.5 mL samples of culture supernatant were collected at 12, 18, 24, and 30 hrs for each transfer. Large molecules were concentrated, from the supernatant using Amicon Ultracel 30K centrifugal filters (Milipore) following the manufacturer's protocol. Thiaminase I assays were then conducted using 3 μ L of the concentrated samples following the protocol described in (66).

Supernatant rescue experiments. Supernatant was collected after centrifugation of E264-*thiC* and BT10432-*thiC* cultures after 24hrs of growth at the transfer before each strain cannot grow (Figure 1). The supernatant was filter through a 0.2 μ m filter. 1.5mL of each supernatant was mixed with 1.5mL of fresh thiamin-free DM4. E264-*thiC* and BT10432 were serially transferred as described in the serial transfer experiments, however, at the terminal transfer where they no longer grow (Figure 1), they were subcultured in the 1:1 E264-*thiC*-DM4:DM4 and 1:1 BT10432-*thiC*-DM4:DM4 mixes.

Thiochrome assay. To ensure that the media was not degrading thiamin we performed a thiochrome assay. Thiamin was added to DM4 at a final concentration of 200 mM and the

production of thiochrome was assayed after being in the media for 24 hr, 12 hr, 1 hr, and 30 minutes. We followed the procedure previously described (67).

Overexpression and purification of thiaminase I. The pET-28 plasmid containing *Clostridium botulinum* thiaminase I (BcmE) from (1) was transformed into *E. coli* BL21 using the protocol in (64). To extract the protein, we modified a previously published protocol (25), with alterations described below. Our procedure used a different lysis buffer (50 mM NaH₂PO₄ and 300 mM NaCl, pH 8.0). 1.5g of USB PrepEase Histidine-Tagged High Yield Purification Resin (Affymetrix) was used for binding the His-tagged BcmE and added to the clear lysate and incubated overnight in an orbital shaker at 4°C. The slurry was then poured into a column and washed with 50 mL of the lysis buffer. The protein was eluted with 15 mL of elution buffer (50mM NaH₂PO₄, 300mM NaCl, and 250mM imidazole, pH 8.0) in increments of 5 mL. Fractions were tested for thiaminase activity (66). The samples were dialyzed overnight in 50 mM potassium phosphate, 100mM NaCl buffer, pH 7.2, and then concentrated using 30,000K Amicon Ultra Centrifugal filters (EMD Milipore) to 1 mg/mL stocks.

Rescue experiments. To test what conditions could recover growth of thiamin auxotrophs purified thiaminase I, thiamin precursors, thiamin, phosphorylated thiamin, or thiamin analogs were added to bacterial cultures. Purified thiaminase I was added to a concentration of 1 µg/mL (n=3). Thiamin and its moieties, HMP (n=6), THZ (n=6), or the combination of both (n=6), thiamin (n=9), TMP (n=6), and TPP (n=6), were added to cultures at 10 µg/mL each. Thiamin analogs, oxythiamin (n=6), pyriethiamin (n=6), 3-benzyl-5-(2-hydroxyethyl)-4-methylthiazolium chloride (n=3), dibenzoyl thiamin (n=3), and thiamin disulfide (n=3) were added at 10µg/mL. For the *thiCthiG*

mutants in both backgrounds, oxythiamin and pyrithiamin were added in concert at a concentration of 10 μ g/mL for each compound.

4.7 REFERENCES

1. Zhang K, Bian J, Deng Y, Smith A, Nunez RE, Li MB, Pal U, Yu A-M, Qiu W, Ealick SE. 2016. Lyme disease spirochaete *Borrelia burgdorferi* does not require thiamin. *Nature microbiology* 2:16213.
2. Kraft CE, Angert ER. 2017. Competition for vitamin B1 (thiamin) structures numerous ecological interactions. *The Quarterly Review of Biology* 92:151-168.
3. Jurgenson CT, Begley TP, Ealick SE. 2009. The structural and biochemical foundations of thiamin biosynthesis. *Annu Rev Biochem* 78:569-603.
4. Schauer K, Stolz J, Scherer S, Fuchs TM. 2009. Both thiamine uptake and biosynthesis of thiamine precursors are required for intracellular replication of *Listeria monocytogenes*. *Journal of bacteriology* 191:2218-2227.
5. Webb E, Claas K, Downs D. 1998. thiBPQ Encodes an ABC Transporter Required for Transport of Thiamine and Thiamine Pyrophosphate in *Salmonella typhimurium*. *Journal of Biological Chemistry* 273:8946-8950.
6. Melnick J, Lis E, Park J-H, Kinsland C, Mori H, Baba T, Perkins J, Schyns G, Vassieva O, Osterman A. 2004. Identification of the two missing bacterial genes involved in thiamine salvage: thiamine pyrophosphokinase and thiamine kinase. *Journal of bacteriology* 186:3660-3662.
7. Karunakaran R, Ebert K, Harvey S, Leonard M, Ramachandran V, Poole P. 2006. Thiamine is synthesized by a salvage pathway in *Rhizobium leguminosarum* bv. *viciae* strain 3841. *Journal of bacteriology* 188:6661-6668.
8. Carini P, Campbell EO, Morré J, Sanudo-Wilhelmy SA, Thrash JC, Bennett SE, Temperton B, Begley T, Giovannoni SJ. 2014. Discovery of a SAR11 growth requirement for thiamin's pyrimidine precursor and its distribution in the Sargasso Sea. *ISME J* 8:1727-1738.
9. Devedjiev Y, Surendranath Y, Derewenda U, Gabrys A, Cooper DR, Zhang R-g, Lezondra L, Joachimiak A, Derewenda ZS. 2004. The structure and ligand binding properties of the *B. subtilis* YkoF gene product, a member of a novel family of thiamin/HMP-binding proteins. *Journal of molecular biology* 343:395-406.
10. Rodionov DA, Hebbeln P, Eudes A, Ter Beek J, Rodionova IA, Erkens GB, Slotboom DJ, Gelfand MS, Osterman AL, Hanson AD. 2009. A novel class of modular transporters for vitamins in prokaryotes. *Journal of bacteriology* 191:42-51.
11. Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS. 2002. Comparative genomics of thiamin biosynthesis in prokaryotes New genes and regulatory mechanisms. *Journal of Biological chemistry* 277:48949-48959.
12. Anderson LN, Koech PK, Plymale AE, Landorf EV, Konopka A, Collart FR, Lipton MS, Romine MF, Wright AT. 2015. Live cell discovery of microbial vitamin transport and enzyme-cofactor interactions. *ACS chemical biology* 11:345-354.
13. Toms AV, Haas AL, Park J-H, Begley TP, Ealick SE. 2005. Structural characterization of the regulatory proteins TenA and TenI from *Bacillus subtilis* and identification of TenA as a thiaminase II. *Biochemistry* 44:2319-2329.

14. Benach J, Edstrom WC, Lee I, Das K, Cooper B, Xiao R, Liu J, Rost B, Acton TB, Montelione GT. 2005. The 2.35 Å structure of the TenA homolog from *Pyrococcus furiosus* supports an enzymatic function in thiamine metabolism. *Acta Crystallographica Section D: Biological Crystallography* 61:589-598.
15. Onozuka M, Konno H, Kawasaki Y, Akaji K, Nosaka K. 2007. Involvement of thiaminase II encoded by the THI20 gene in thiamin salvage of *Saccharomyces cerevisiae*. *FEMS yeast research* 8:266-275.
16. Zallot R, Yazdani M, Goyer A, Ziemak MJ, Guan J-C, McCarty DR, de Crécy-Lagard V, Gerdes S, Garrett TJ, Benach J. 2014. Salvage of the thiamin pyrimidine moiety by plant TenA proteins lacking an active-site cysteine. *Biochemical Journal* 463:145-155.
17. Jenkins AH, Schyns G, Potot S, Sun G, Begley TP. 2007. A new thiamin salvage pathway. *Nature chemical biology* 3:492-497.
18. Kreinbring CA, Remillard SP, Hubbard P, Brodtkin HR, Leeper FJ, Hawksley D, Lai EY, Fulton C, Petsko GA, Ringe D. 2014. Structure of a eukaryotic thiaminase I. *Proceedings of the National Academy of Sciences* 111:137-142.
19. Costello CA, Kelleher NL, Abe M, McLafferty FW, Begley TP. 1996. Mechanistic studies on thiaminase I overexpression and identification of the active site nucleophile. *Journal of Biological Chemistry* 271:3445-3452.
20. Fujita A. 1954. Thiaminase. *Advances in Enzymology and Related Areas of Molecular Biology*, Volume 15:389-421.
21. Agee CC, Airth R. 1973. Reversible inactivation of thiaminase I of *Bacillus thiaminolyticus* by its primary substrate, thiamine. *Journal of bacteriology* 115:957-965.
22. Ebata J, Murata K. 1961. The purification of thiaminase I produced by *Bacillus thiaminolyticus*. *The Journal of vitaminology* 7:115-121.
23. Wang L, Airth R. 1967. Repression of thiaminase I in *Bacillus thiaminolyticus*. *Biochemical and biophysical research communications* 27:325-330.
24. Wang L, Wilkins JH, Airth R. 1968. Repression of thiaminase I by thiamine and related compounds in *Bacillus thiaminolyticus*. *Canadian journal of microbiology* 14:1143-1147.
25. Sikowitz MD, Shome B, Zhang Y, Begley TP, Ealick SE. 2013. Structure of a *Clostridium botulinum* C143S thiaminase I/thiamin complex reveals active site architecture. *Biochemistry* 52.
26. Soriano EV, Rajashankar KR, Hanes JW, Bale S, Begley TP, Ealick SE. 2008. Structural similarities between thiamin-binding protein and thiaminase-I suggest a common ancestor. *Biochemistry* 47:1346-1357.
27. Cooper LE, O'Leary SnE, Begley TP. 2014. Biosynthesis of a Thiamin Antivitamin in *Clostridium botulinum*. *Biochemistry* 53:2215-2217.
28. Green R, Evans C. 1940. A deficiency disease of foxes. *Science* 92:154-155.
29. Evans C, Carlosn W, Green R. 1942. The Pathology of Chastek Paralysis in Foxes: A Counterpart of Wernicke's Hemorrhagic Polioencephalitis of Man. *The American journal of pathology* 18:79.
30. Yudkin WH. 1949. Thiaminase, the Chastek-paralysis factor. *Physiological reviews* 29:389-402.
31. Linklater K, Dyson D, Morgan K. 1977. Faecal thiaminase in clinically normal sheep associated with outbreaks of polioencephalomalacia. *Research in veterinary science* 22:308-312.

32. Shreeve J, Edwin E. 1974. Thiaminase-producing strains of *Cl. Sporogenes* associated with outbreaks of cerebrocortical necrosis. *The Veterinary record* 94:330.
33. Honeyfield DC, Ross JP, Carbonneau DA, Terrell SP, Woodward AR, Schoeb TR, Perceval HF, Hinterkopf JP. 2008. Pathology, physiologic parameters, tissue contaminants, and tissue thiamine in morbid and healthy central Florida adult American alligators (*Alligator mississippiensis*). *Journal of wildlife diseases* 44:280-294.
34. Honeyfield DC, Hinterkopf JP, Fitzsimons JD, Tillitt DE, Zajicek JL, Brown SB. 2005. Development of thiamine deficiencies and early mortality syndrome in lake trout by feeding experimental and feral fish diets containing thiaminase. *J Aquat Anim Health* 17:4-12.
35. Brown SB, Fitzsimons JD, Honeyfield DC, Tillitt DE. 2005. Implications of thiamine deficiency in Great Lakes salmonines. *Journal of Aquatic Animal Health* 17:113-124.
36. Honeyfield DC, Hinterkopf JP, Brown SB. 2002. Isolation of thiaminase-positive bacteria from alewife. *Transactions of the American Fisheries Society* 131:171-175.
37. Richter CA, Evans AN, Wright-Osment MK, Zajicek JL, Heppell SA, Riley SC, Krueger CC, Tillitt DE. 2012. *Paenibacillus thiaminolyticus* is not the cause of thiamine deficiency impeding lake trout (*Salvelinus namaycush*) recruitment in the Great Lakes. *Canadian Journal of Fisheries and Aquatic Sciences* 69:1056-1064.
38. Nishimune T, Watanabe Y, Okazaki H, Akai H. 2000. Thiamin is decomposed due to *Anophe* spp. entomophagy in seasonal ataxia patients in Nigeria. *The Journal of nutrition* 130:1625-1628.
39. Erkens GB, Slotboom DJ. 2010. Biochemical characterization of ThiT from *Lactococcus lactis*: a thiamin transporter with picomolar substrate binding affinity. *Biochemistry* 49:3203-3212.
40. Douthit H, Airth R. 1966. Thiaminase I of *Bacillus thiaminolyticus*. *Archives of biochemistry and biophysics* 113:331-337.
41. Murata K. 1982. Actions of two types of thiaminase on thiamin and its analogues. *Annals of the New York Academy of Sciences* 378:146-156.
42. Shin W, Chun K. 1987. Structure of thiamin disulfide dinitrate. *Acta Crystallographica Section C: Crystal Structure Communications* 43:2123-2125.
43. Fukui S, Ohishi N, Kishimoto S, Takamizawa A, Hamazima Y. 1965. Formation of "Thiaminosuccinic Acid" as an Intermediate in the Transformation of Oxythiamine to Thiamine by a Thiamineless Mutant of *Escherichia coli*. *Journal of Biological Chemistry* 240:1315-1321.
44. Schmidhalter U, Evéquo M, Studer C, Oertli J, Kahr G. 1994. Adsorption of thiamin (vitamin B1) on soils and clays. *Soil Science Society of America Journal* 58:1829-1837.
45. Cain AH, Sullivan GR, Roberts JD. 1977. The protonation site of vitamin B1 as determined from natural-abundance nitrogen-15 nuclear magnetic resonance spectra. *Journal of the American Chemical Society* 99:6423-6425.
46. Dwivedi BK, Arnold RG. 1973. Chemistry of thiamine degradation on food products and model systems. Review. *Journal of Agricultural and Food Chemistry* 21:54-60.
47. Linster CL, Van Schaftingen E, Hanson AD. 2013. Metabolite damage and its repair or pre-emption. *Nature chemical biology* 9:72-80.
48. Goyer A, Hasnain G, Frelin O, Ralat MA, Gregory JF, Hanson AD. 2013. A cross-kingdom Nudix enzyme that pre-empts damage in thiamin metabolism. *Biochemical Journal* 454:533-542.

49. Pribat A, Blaby IK, Lara-Núñez A, Jeanguenin L, Fouquet R, Frelin O, Gregory JF, Philmus B, Begley TP, de Crécy-Lagard V. 2011. A 5-formyltetrahydrofolate cycloligase paralog from all domains of life: comparative genomic and experimental evidence for a cryptic role in thiamin metabolism. *Functional & integrative genomics* 11:467-478.
50. Schwarz S, West TE, Boyer F, Chiang W-C, Carl MA, Hood RD, Rohmer L, Tolker-Nielsen T, Skerrett SJ, Mougous JD. 2010. Burkholderia type VI secretion systems have distinct roles in eukaryotic and bacterial cell interactions. *PLoS pathogens* 6:e1001068.
51. DeShazer D. 2007. Virulence of clinical and environmental isolates of Burkholderia oklahomensis and Burkholderia thailandensis in hamsters and mice. *FEMS microbiology letters* 277:64-69.
52. Haraga A, West TE, Brittnacher MJ, Skerrett SJ, Miller SI. 2008. Burkholderia thailandensis as a model system for the study of the virulence-associated type III secretion system of Burkholderia pseudomallei. *Infection and immunity* 76:5402-5411.
53. Pilátová M, Dionne MS. 2012. Burkholderia thailandensis is virulent in Drosophila melanogaster. *PLoS One* 7:e49745.
54. Fisher NA, Ribot WJ, Applefeld W, DeShazer D. 2012. The Madagascar hissing cockroach as a novel surrogate host for Burkholderia pseudomallei, B. mallei and B. thailandensis. *BMC microbiology* 12:117.
55. Haughton BG, King H. 1958. Toxopyrimidine phosphate as an inhibitor of bacterial enzyme systems that require pyridoxal phosphate. *Biochemical Journal* 70:660.
56. Shanks RM, Kadouri DE, MacEachran DP, O'Toole GA. 2009. New yeast recombineering tools for bacteria. *Plasmid* 62:88-97.
57. Neidhardt FC, Bloch PL, Smith DF. 1974. Culture medium for enterobacteria. *J Bacteriol* 119:736-747.
58. Harwood CR, Cutting SM. 1990. Molecular biological methods for Bacillus. Wiley.
59. Farrer K, Hollenberg W. 1953. Adsorption of thiamine on glassware. *Analyst* 78:730-731.
60. Brett PJ, DeShazer D, Woods DE. 1998. Burkholderia thailandensis sp. nov., a Burkholderia pseudomallei-like species. *International journal of systematic bacteriology* 48 Pt 1:317-320.
61. Gallagher LA, Ramage E, Patrapuvich R, Weiss E, Brittnacher M, Manoil C. 2013. Sequence-defined transposon mutant library of Burkholderia thailandensis. *MBio* 4:e00604-13.
62. Shanks RM, Caiazza NC, Hinsa SM, Toutain CM, O'Toole GA. 2006. Saccharomyces cerevisiae-based molecular tool kit for manipulation of genes from gram-negative bacteria. *Applied and environmental microbiology* 72:5027-5036.
63. Burke D, Dawson D, Stearns T. 2000. Methods in yeast genetics: a Cold Spring Harbor Laboratory course manual. CSHL Press.
64. Peters JE. 2007. Gene transfer in Gram-negative bacteria, p 735-755, *Methods for General and Molecular Microbiology*, Third Edition. American Society of Microbiology.
65. Thongdee M, Gallagher LA, Schell M, Dharakul T, Songsivilai S, Manoil C. 2008. Targeted mutagenesis of Burkholderia thailandensis and Burkholderia pseudomallei through natural transformation of PCR fragments. *Applied and environmental microbiology* 74:2985-2989.
66. Kraft CE, Gordon ER, Angert ER. 2014. A Rapid Method for Assaying Thiaminase I Activity in Diverse Biological Samples. *PloS one* 9:e92688.

67. Edwards KA, Seog WJ, Han L, Feder S, Kraft CE, Baeumner AJ. 2016. High-Throughput Detection of Thiamine Using Periplasmic Binding Protein-Based Biorecognition. Anal Chem 88:8248-8256.

4.8 SUPPLEMENTARY MATERIAL

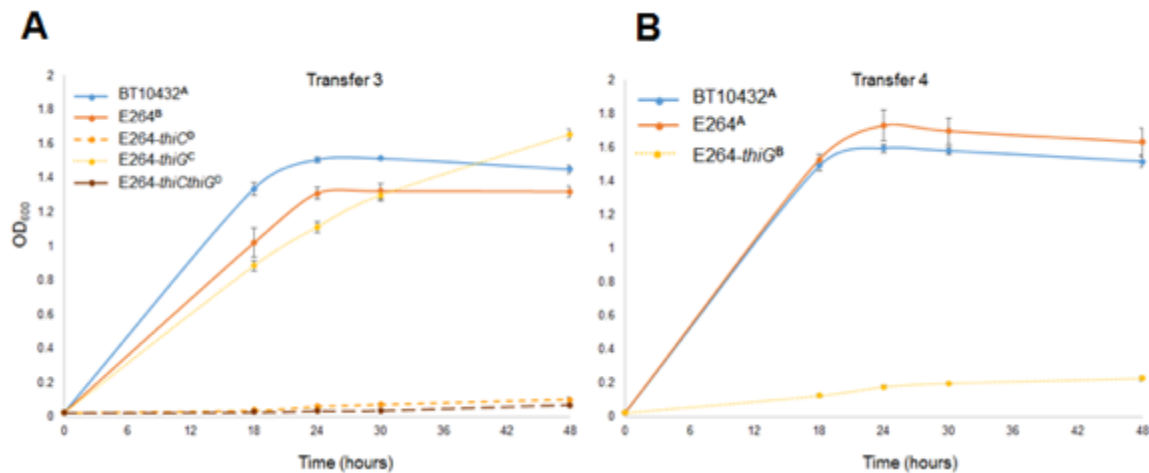


FIG S1 Growth of E264 and BT10432 strains in defined media serial transfers. (A) Indicates the growth of the remaining strains in the third transfer in DM4. The E264-*thiC* and *thiCthiG* mutants were unable to grow. BT10432 reached a significantly higher OD₆₀₀ at 24rs ($p < 0.0001$ for all significantly different pairwise comparisons) than the other 2 strains. E264-*thiG* was significantly higher than E264 ($p < 0.0001$). (B) Indicates the growth of the 3 strains in the fourth transfer in DM4. There is no statistically significant difference between BT10432 and E264 growth. The growth curves displayed are the average of three replicate cultures with standard error bars at each time point. In each case, growth at 24 hr was compared and significant differences are designated by different superscripts next to the names in the legend.

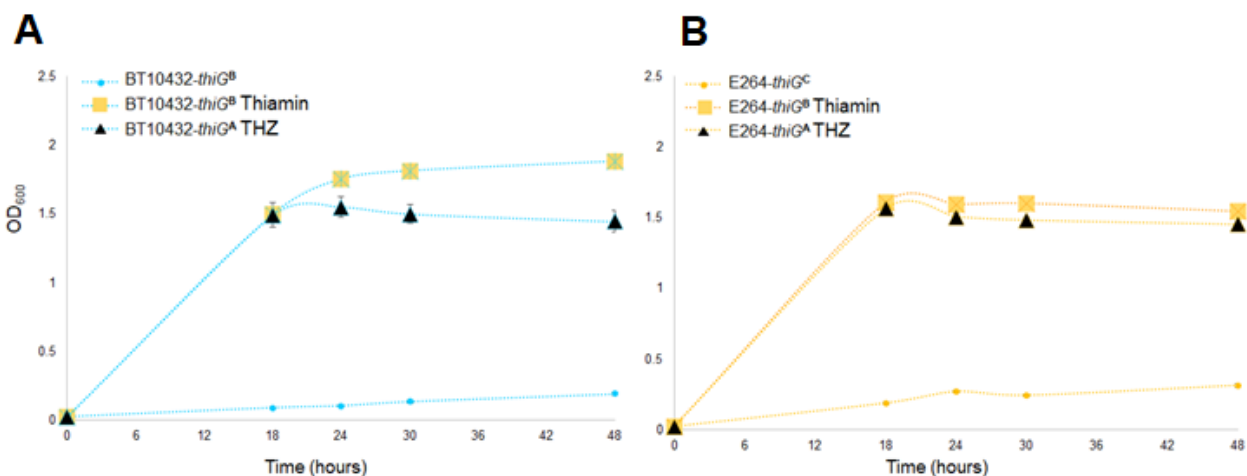


FIG S2 Growth of *B. thailandensis* *thiG* strains on thiamin and THZ. (A) BT10432-*thiG* reaches a significantly higher OD600 at 24hrs ($p = 0.0012$) when rescued with thiamin in contrast to THZ. (B) There was no difference in growth when the E264-*thiG* strain was grown on THZ or thiamin. The growth curves displayed are the average of six replicate cultures with standard error bars at each time point. In each case, growth at 24 hr was compared and significant differences are designated by different superscripts next to the names in the legend.

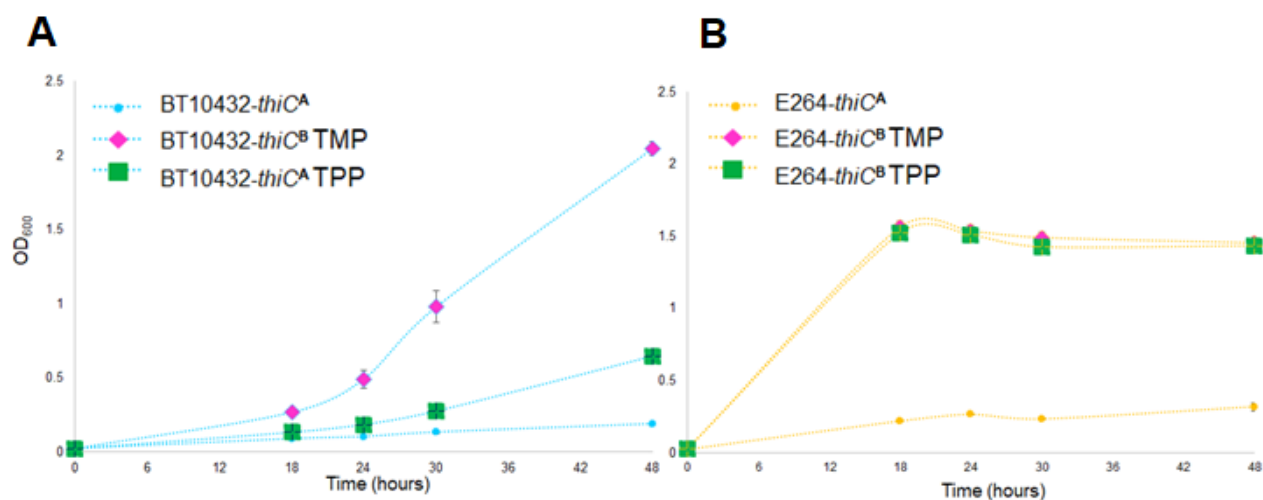


FIG S3 Growth of *B. thailandensis* *thiG* strains on TMP and TPP. (A) shows there is no substantial growth of BT10432-*thiG* when TMP is present and this is significantly higher than the moderate growth observed when TPP is present ($p < 0.0001$). (B) shows that E264-*thiG* was able to use both TMP and TPP to support its auxotrophy. The growth curves displayed are the average of six replicate cultures with standard error bars at each time point. In each case, growth at 24 hr was compared and significant differences are designated by different superscripts next to the names in the legend.

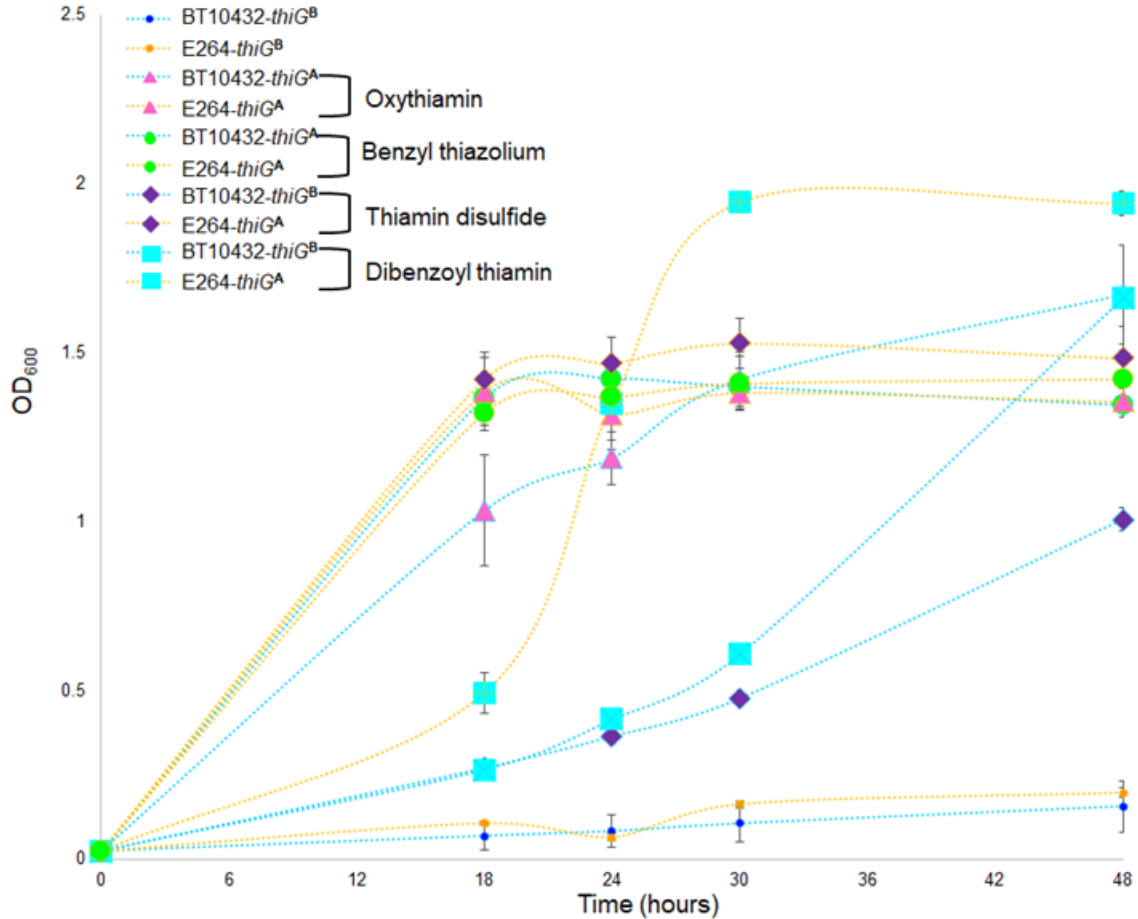


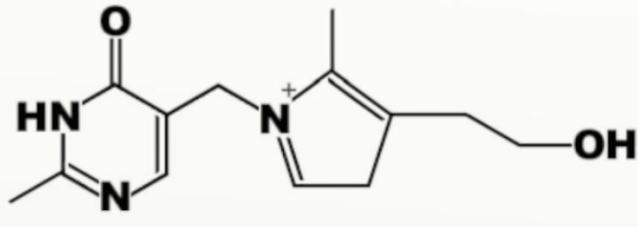
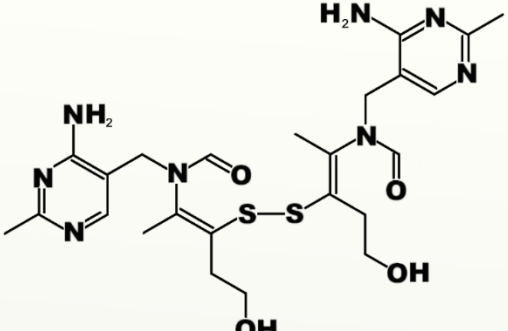
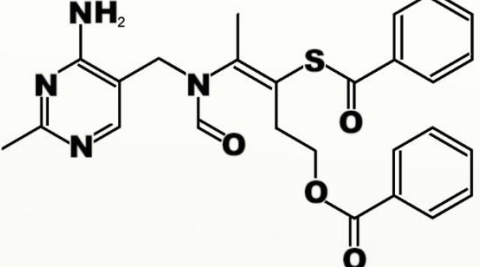
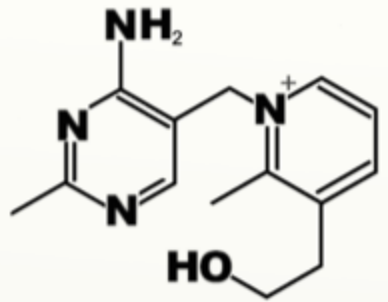
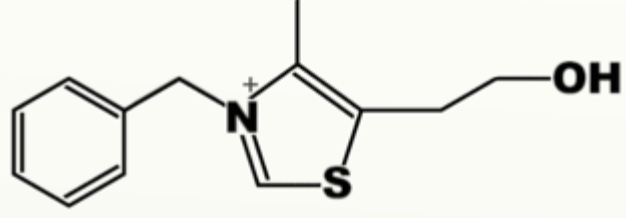
FIG S4 Growth of *B. thailandensis* *thiG* strains on thiamin analogs. The graph shows that both BT10432-*thiG* and E264-*thiG* are able to grow on oxythiamin and 3-benzyl-5-(2-hydroxyethyl)-4-methylthiazolium chloride, abbreviated as benzyl thiazolium in the figure legend. For all significant differences in pairwise comparisons, $p < 0.0001$. Unlike BT10432-*thiG*, E264-*thiG* can grow well on dibenzoyl thiamin, with a significantly higher OD₆₀₀ than BT10432-*thiG*. At 24hr E264-*thiG* had significantly higher growth than BT10432-*thiG* on thiamin disulfide, though BT10432-*thiG* grew slowly on this analog. The growth curves displayed are the average of three replicate cultures with standard error bars at each time point. In each case, growth at 24 hr was compared and significant differences are designated by different superscripts next to the names in the legend.

Table S1 Primer list used in this study

Primer name	Sequence
pMQ87- <i>thiC</i> 5' F	<u>AGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCTCTGTCCC</u> <u>CTGTTGAAAC</u>
<i>thiC</i> 5' F	TCTGTCCCCTGTTGAAAC
<i>thiC</i> 5' R	CAGGCTTTCCAGATACTC
<i>thiC</i> 3' F	GACTGGTGCAAGGAAGCG
<i>thiC</i> 3' R	TTACGCTGCTGGGTGGTC
pMQ87- <i>thiC</i> 3' R	<u>AAACAGCTATGACCATGATTACGAATTCGAGCTCGTTACGCT</u> <u>GCTGGGTGGTC</u>
<i>thiC-tetR</i> F	CGCGATCCGCGAGAACCAGCGCCGCGCCGAGTATCTGGA AAGGCTAACGCAGTCAGGCAC
<i>tetR-thiC</i> R	GTCGTAGCCGGGCGCGATGTCGGTCGTGAGCGGCCCCGAG CGTTCCGTTAGCGAGGTGCCG
<i>thiC-tetR</i> construct confirmation F	GCGCAACATGGCATCCGG
<i>thiC-tetR</i> construct confirmation R	CGTCGGCCCATTCGAGGG
pMQ87- <i>thiG</i> 5' F	<u>AGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCC</u> <u>AGCTGCA</u> <u>CGGTCCGCAC</u>
<i>thiG</i> 5' F	AGCTGCACGGTCCGCAC
<i>thiG</i> 5' R	GCGTGCTGCTGAACACG
<i>thiG</i> 3' F	AGCGCGACCGTCACCATC
<i>thiG</i> 3' R	GAAGTGAATGCGCAATGC
pMQ87- <i>thiG</i> 3' R	<u>AAACAGCTATGACCATGATTACGAATTCGAGCTCG</u> <u>GAACTGA</u> <u>ATGCGCAATGC</u>
<i>thiG-cat</i> F	GAACGCGCGCGCCATGATCTCCGGATGCGTCGCCTGCGA TACGGCGAAAATGAGACGTTG
<i>cat-thiG</i> R	GTCGCTGCAGTCGCTGTCCGATTCGATCGCGGCCGCGCG GCCAGGCGTAGCAACCAGGCG
<i>thiG-cat</i> construct confirmation F	GTTCAGGTTTCGACAGGAC
<i>thiG-cat</i> construct confirmation R	ATGCGATGAACATGCGTG

Underlines signify the section of the primers that overlap with pMQ87. Bold font represents sections of the antibiotic resistance gene primers that overlap with sections of thiamin biosynthesis genes

Table S2 Thiamin analog structures

Name	Structure
Oxythiamin	
Thiamin disulfide	
Dibenzoyl thiamin	
Pyrithiamin	
3-benzyl-5-(2-hydroxyethyl)-4-methylthiazolium	

CHAPTER 5

INVESTIGATIONS INTO THE METABOLIC AND BIOSYNTHETIC POTENTIAL OF ‘CANDIDATUS EPULOPISCIMUM VIVIPAROUS’

5.1 ABSTRACT

Epulopiscium spp. are the largest known heterotrophic bacteria. They can exceed cellular lengths of 600 μm and are gut residents of tropical marine herbivorous surgeonfish, but this symbiotic association is not well defined. The best characterized member of the *Epulopiscium* group is referred to simply as *Epulopiscium* sp. type B. To better understand its metabolic and biosynthetic potential as well as its relationship with its host *Naso tonganus*, we generated a high quality draft genome from a population of type B cells, and here propose the name ‘*Candidatus Epulopiscium viviparous*’. Genomic analysis of their 3.28 Mb chromosome confirms their unique mode of reproduction through multiple internal offspring formation is derived from endosporulation. The genome is enriched for carbohydrate metabolism and transport as Cluster of Orthologous Genes analysis revealed they dedicated the highest proportion of genes to this functional group, and greater than 5% of the genome encodes for carbohydrate active enzymes. ‘*Ca. E. viviparous*’ has the capacity to degrade multiple recalcitrant algal polysaccharides, which can be further fermented to acetate for host utilization. To further conserve energy, ‘*Ca. E. viviparous*’ can generate ATP through a sodium-motive-force, which is likely produced by the decarboxylation of oxaloacetate and nitrate is used as an electron sink for redox balancing. *Ca. E. viviparous* has the potential to assimilate host nitrogenous waste products into protein as it has the ability to synthesize all 20 amino acids. This protein and other organic nitrogen produced by the bacterium may supplement the nutritional needs of its host. The analysis provides insight into the exceptional physiology of this unusual heterotroph.

5.2 INTRODUCTION

As in mammals, the gastrointestinal (GI) tracts of fish harbor complex, specialized microbial communities (1). Community composition is dynamic and likely shaped by numerous biotic factors including host diet (2-4), genotype (5), phylogeny (3), developmental stage (6, 7), gut physiology (1), sex (8) and location within (gut lumen versus mucosal layer) (9) and along the GI tract (6). Abiotic factors such as ambient temperature, season, and salinity also impact community structure (6, 10). The intestinal microbiota affect their host in a variety of ways. Gnotobiotic studies in zebrafish reveal that host health is impacted through the microbial stimulation of epithelial cells lining the intestinal tract which stimulates differentiation, maturation, and cell proliferation (11-13). Resident microbial populations prevent the colonization by pathogens, and induce development of the adaptive immune system (14). In other fish species GI bacteria have been shown to aid in host nutrition by producing nutrients such as B vitamins (15), degradative enzymes such as phosphatases, chitinases, proteases, and lipases (6, 16), and aid protein macromolecule uptake (12), as well as cholesterol trafficking (11). However, due to the paucity of *in vivo* functional analyses, our understanding of fish gastrointestinal microbiota is still lacking in comparison to that of terrestrial vertebrate GI communities (1).

The hindgut microbiota of some marine herbivorous and omnivorous fish appear to aid in digestion. Complex polysaccharides such as carrageenan, laminarin, agar, and alginic acid, and sugar alcohols like mannitol are constituents of the algal diets of some of these fish, but these compounds are not readily assimilated (17). It is likely that gut symbionts produce enzymes that breakdown polysaccharides into fermentable products which are converted into short-chain fatty acids (SCFA) mainly in the posterior regions of the intestine (17, 18). The SCFA can be absorbed by the intestines and contribute energy to the host, however, a comprehensive understanding of

how the symbionts fully influence host nutrition is lacking (17, 19). Unlike other fish whose gut communities are dominated by Proteobacteria, the guts of the marine herbivorous fish studied thus far resemble the intestinal microbiota of mammals, as members of the Firmicutes and the Bacteroidetes predominate these communities (10). Clements and colleagues demonstrated that members of clostridial clusters XI and XIVa (Lachnospiraciae) dominate the guts of temperate marine herbivores (20). Due to repertoire of degradative enzymes found in many gut associated Firmicutes and Bacteroidetes, it seems likely that they are degrading and fermenting the recalcitrant polysaccharides present in the host diets.

The gut communities of herbivorous and omnivorous members of tropical coral reef dwelling surgeonfish of the family Acanthuridae are mainly dominated by Firmicutes, in particular *Epulopiscium* spp. (and related bacteria referred to as 'epulos') (1-3, 21). *Epulopiscium* spp. are members of the Lachnospiraceae, falling in the XIVb cluster of the Clostridiales (22, 23), and are rather unique in the known bacterial world. *Epulopiscium* spp. boast the largest known heterotrophic bacteria, and are discriminated from one another and categorized into morphotypes based on their size, shape (cigar, ellipsoid, or elongate), and reproductive strategy (2, 24). The cigar shaped type A, was the first discovered, found in the GI tract of surgeonfish from the Red Sea (25) and later found in surgeonfish in the South Pacific. Cells can reach lengths exceeding 600 μm and widths of 80 μm , with cytoplasmic volumes of 2,000,000 μm^3 (24, 26). Reproductive strategies of epulos include relying solely on the production of multiple, internal offspring (type A and B) or dormant endospores (type C and J) for propagation, or binary fission (types G, I, and J) (2, 21).

Type B is the best characterized of the epulo morphotypes. Unlike other epulos, which are found in multiple hosts in concert with one another, type B is found in adult *Naso tonganus*, where

it is the only epulo present in the GI community. Type B populations are morphologically and genetically more homogenous than other morphotypes (2, 27). Cells reach lengths of 300 μm and widths of 60 μm and contain massive amounts of DNA, with large type B cells containing approximately 250 pg of DNA (28). The DNA is situated at the periphery of the cytoplasm in nucleoids that are arranged in a mesh-like network (29). Further, Mendell et al. determined that type B cells are extremely polyploid, containing tens of thousands of copies of their chromosomes, with chromosomal copy number proportional to cell size (28). Extreme polyploidy and peripheral arrangement of DNA may be size adaptations that negate the constraints of having a low surface area to volume ratio and potentially allow for local responses to stimuli at different parts of the cell (28).

Another distinguishing feature of type B is their ability to produce up to twelve intracellular offspring in a developmental program that resembles endospore formation found in other Firmicutes (30) and the mode of propagation found in the Guinea pig symbiont *Metabacterium polyspora* (31). Hallmarks shared between these processes include mother-cell DNA coalescing at the poles prior to offspring formation, asymmetric bipolar division, and engulfment of the offspring (23). In type B, it appears that approximately 1% of the total mother cell DNA, is translocated into each daughter cell. The daughter cells develop inside the mother cell, while they replicate their genomes (32), and eventually fill the mother cell's cytoplasm (23, 27). When the daughter cells fully mature, the mother cell lyses, releasing the daughter cell (23). This life cycle perpetuates in a diurnal fashion where cells are synchronized in regards to life stages, and this corresponds to host feeding activity (23). This evolutionary framework is further supported by genomic evidence, as Miller and colleagues demonstrated that genes for the central regulatory network of sporulation are conserved in *Epulopiscium* sp. type B (27).

Our understanding of *Epulopiscium* biology is still limited. *Epulopiscium* spp. are not yet in culture, their metabolism is not understood, and their symbiotic association with their surgeonfish host is not well characterized. Here we report a high quality draft of the circular 3.82 Mb genome of *Epulopiscium* sp. type B, for which we propose the name ‘*Candidatus* *Epulopiscium viviparous*’ due to its extreme ability to produce up to 12 internal offspring (though it will still be referred to as type B throughout the text for convenience). The genome corroborates findings that this extreme mode of reproduction is derived from endosporulation but the capacity to form dormant spores has been lost. Metabolic reconstructions based on genomic evidence assert that type B is an obligate fermenter whose genome is primed for carbohydrate metabolism. It encodes for a diverse array of carbohydrate active enzymes to utilize a broad spectrum of polysaccharides likely present in the host diet. These polysaccharides can be fermented to products *N. tonganus* can assimilate. Further, the genome provides ‘*Ca. E. viviparous*’ with the ability to conserve energy through ATP synthesis. The cell does not rely solely on substrate level phosphorylation but rather it can drive its ATPase using a sodium motive force generated through oxaloacetate decarboxylation. The genome also provides type B with the ability to incorporate nitrogenous waste products from the host into biomass. This in turn potentially benefits its host as ‘*Ca. E. viviparous*’ releases some of its biomass when offspring cells are released. Similarly, type B can produce multiple B vitamins, giving it the potential to serve as a nutritional reservoir for *N. tonganus*. The genome paints a broad picture of ‘*Ca. E. viviparous*’ biology, allowing us to better understand the metabolic and biosynthetic capabilities of this organism and how they relate to its unique physiology and association with *N. tonganus*.

5.3 MATERIALS AND METHODS

Sample collection

Naso tonganus Nt_450 was caught off of Lizard Island, Australia in May 2013 via spearfishing. It was dissected, its guts were removed, and contents from the gut were fixed in 80% ethanol and stored at -20°C.

DNA extraction

Large chunks of debris were settled out of the fixed gut contents and the supernatant containing '*Ca* Epulopiscium viviparous' cells was transferred to a sterile 1.5 mL Eppendorf tube. There, the type B cells were settled out and supernatant was removed to eliminate smaller microbes. The type B cells were washed two times with sterile 80% ethanol wash buffer containing 145 mM NaCl, 50 mM Tris-Cl pH 8.0, and 0.05% Ipegal. 35,000 clean individual '*Ca* Epulopiscium viviparous' cells were hand picked using a standard Gilson P20 pipettor and a Nikon SMZ-U dissecting microscope, and transferred to sterile 80% ethanol wash buffer.

The 35,000 cells were pelleted and resuspended in lysis buffer containing 10 mM Tris pH 8, 5 mM EDTA, and 0.5% SDS, and subject to degradation with 400 µg/mL proteinase K overnight. DNA was extracted with phenol:chloroform:isoamyl alcohol, buffered to pH 8.0, precipitated with 100% ethanol, and treated with RNase A. Genomic DNA was visualized on a 1% agarose gel via electrophoresis and total genomic DNA was quantified using a Qubit 3.0 fluorometer (ThermoFisher Scientific). As a quality control measure, the gene for the 16S rRNA was PCR amplified using universal primers 515F and 1492R (33) from the DNA extraction, and sequenced using Sanger sequencing at the Cornell University Institute of Biotechnology to ensure only the '*Ca* Epulopiscium viviparous' 16S rRNA gene was amplified.

Whole genome sequencing and assembly

For PacBio sequencing, high molecular weight genomic DNA was submitted to the Yale Center for Genome Analysis, and a 10-20 kb library prep was generated, though the actual average fragment size recovered was 7 kb. The library was sequenced using PacBio RSII in 1 SMRT cell with P6-C4 chemistry. The DNA was also submitted for analysis to the Cornell University Biotechnology Resource Center for 2 x 250 bp paired-end Illumina MiSeq platform.

Multiple genome assemblers were used to generate the '*Ca* Epulopiscium viviparous' draft genome. HGAP assembler (34) and Canu 1.1 (35) were used separately for PacBio only assembly in which reads were cleaned and self-corrected, and assembled. HGAP assemblies were polished using Quiver. Corrected reads from the best Canu assembly were then assembled using MIRA 4.9.5 (36), SPAdes 3.7.1 (37), and dipSPAdes 1.0 (38). Illumina only assemblies were also conducted. Illumina reads were all processed using the BBtools software package, in which they were merged, error corrected and normalized and used as input for MIRA 4.9.5, and SPAdes 3.7.1. Raw Illumina reads were used as input for metaSPAdes (39), or as input for MIRA 4.9.5 and SPAdes 3.7.1, where the assemblers used their native software to error-correct and join reads. Hybrid assemblies were conducted using both PacBio and Illumina data in SPAdes 3.7.1, MIRA 4.9.5, and dipSPAdes 1.0. The hybrid assemblies utilized either the BBtools processed Illumina reads and Canu self-corrected PacBio reads, or the raw Illumina and Canu self-corrected PacBio reads. Another approach tried was correcting raw PacBio reads with Illumina reads using LoRDEC (40), followed by assembly using Canu 1.1 or SPAdes 3.7.1. All the assemblies were compared with each other based on total numbers of contigs, genome length, G+C content, and N50.

Promising assemblies with low contig numbers and reasonable genome (assemblies that fit within the range of our previously assembled genome and slightly exceeding the length predicted to be 3.8Mb (28)) were further analyzed and refined. Read coverage was determined by mapping back the trimmed Illumina reads using bbmap and samtools. BLASR (41) was used for determining the read coverage of the PacBio reads. For these assemblies, contigs were sorted by coverage and G+C content. Short, low-coverage contigs outside the appropriate GC% range were discarded from the assemblies. Assembly networks using linkage mapping based on overlapping read coverage of the paired-end Illumina reads were generated using BamM, allowing for further refinement of the assemblies. Upon refinement, assemblies were assessed for completeness, heterogeneity, and contamination using CheckM (42) at the phylum level, as well as through a phylogenetically targeted conserved list of single-copy genes. The genes present in this list include those experimentally validated to be essential for *Bacillus subtilis* and considered single-copy in 95% of all bacterial genomes according to the Comprehensive Microbial Resource database (43), and predicted as the core minimal bacterial gene set required for life in the absence of environmental stress (44). The genes were normalized and validated against the complete genome of *Cellulosyliticum lentocellum* DSM 5427, the closest free-living, sequenced relative of type B. Duplications were determined from both CheckM and the gene list, and contigs were manually curated and removed if duplicated.

To close gaps in the best assembly (the LoRDEC-Canu assembly), the other assemblies were utilized and aligned against it using Mauve. If multiple assemblies from different assemblers shared the same open reading frames and spanned gaps, they were incorporated into the assembly, to close gaps. PCR was attempted to close remaining gaps that could not be closed through this method.

Terminus prediction

We attempted to identify the terminus of the near complete genome of ‘*Ca. Epulopiscium viviparous* str. B M450’ with the visual aid of the GC skew outputs from DNAPlotter (45) and the GC Skew program in iRep (<https://github.com/christophertbrown/iRep>) (46). The closed genome of *C. lentocellum* DSM 5427 was used as a reference for GC skew visualization and a *xerC* homolog was located.

Genome annotation and bioinformatics

The refined LoRDEC-Canu genome was submitted to IMG (47) for annotation using the MGAP v4 pipeline (48), as well as to Genoscope for annotation through the Microscope platform (49). Metabolic model of ‘*Ca. E. viviparous*’ was constructed using the KEGG maps generated by both platforms, as well as through Microcyc pathways. If genes were missing in certain pathways, the *C. lentocellum* DSM 5427 genome was interrogated to locate a homolog of the missing gene. If found, the predicted amino acid sequence was queried against the ‘*Ca. E. viviparous*’ genome using the local BLAST in Genoscope. If *C. lentocellum* DSM 5247 was missing the gene, amino acid sequences of homologs from other Lachnospiraceae, Clostridia, or other Firmicutes from the Uniprot database were queried against the genome. The same method was applied for elucidating pathways found in other organisms performing relevant functions after searching the literature. Candidates for the missing genes were analyzed via DELTA-BLAST and the RCSB protein databank (50) to ensure that they contained the proper domains found in the protein missing from the pathways. Gene synteny, using *C. lentocellum* DSM 5427 and *B. subtilis* 168 genomes as references, was also used to identify the function of poorly annotated genes. To identify Carbohydrate Active enZymes, searches were performed using the HmmerScan program in HMMER (3.1) against the CAZyme family-specific HMM (hidden Markov model) database

downloaded from dbCAN (51). Results were manually parsed with an E-value cutoff 10^{-5} and the genomic context of the CAZymes was examined for biological relevance. Overlapping domains were removed manually from the proteins. IMG was used to determine secretory proteins, which were verified using SignalP 4.1(52). To determine the substrates for the carbohydrate transport proteins, the amino acid sequences were queried against the Transporter Classification Database (53), as well as through DELTA-BLAST and the RCSB protein database. For all BLAST searches in the various databases, we used an E-value cutoff of $<10^{-5}$, minimal coverage $>50\%$. To examine the domains of the proteins from large genes over 5 kb in length, we used InterProScan (54). Tandem repeats were found in the genome using Tandem Repeat Finder (55). Repeats were filtered for having a period length of ≥ 9 bp with 10 or more occurrences.

Cluster of Orthologous Genes analysis

We performed a Cluster of Orthologous Genes (COG) analysis on the type B genome, as well as on other draft or complete genomes of relatives. We used [*Clostridium*] *neopropionicum* X4 DSM 3847, [*Clostridium*] *propionicum* DSM 1682, *C. lentocellum* DSM 5427, *Cellulosilyticum* sp. I15G10I2, *C. ruminicola* JCM 14822, and *Niameybacter massiliensis* str. MT14 as members of the Lachnospiraceae XIVb cluster for comparison. *Butyrivibrio crossotus* DSM 2876, *Coprococcus* sp. HPP0074, and *Roseburia intestinalis* L1-82 were used as representatives of the Lachnospiraceae XIVa cluster. Prodigal 2.6.2 (56) was used to generate ORFs and protein sequences for each genome and these protein sequences were subsequently BLASTed against the COG database. Sequences with best hits and E-value $> 10^{-5}$ were assigned unique matches to COG clusters after parsing in R. The proportion of genes assigned to each COG

category was calculated, but some gene queries can be classified into multiple COGs overestimating the total gene counts per genome.

5.4 RESULTS

Draft genome features of ‘*Ca. Epulopiscium viviparous*’

Table 1. Statistics of the ‘*Ca. Epulopiscium viviparous*’ genome.

Genome feature	Value
Genome size (bp)	3,282,201
Contigs	7
Max contig length	905,387
N50 (bp)	602,975
G+C content %	38.08%
Coding percentage	91.96%
ORFs	2,714
Protein coding genes	2,635
tRNAs	54
16S rRNA genes	6
23S rRNA genes	6
5S rRNA genes	6
Genome completeness based on CheckM (%)	97%
Genome completeness based on conserved gene list (%)	92%

Through an iterative hybrid assembly approach, using PacBio reads corrected with Illumina reads, we were able to generate a high quality draft genome according to the criteria set forth by Bowers and colleagues (57), assembled into 7 contigs ranging from almost 1 Mb to 30 kb, with all but 1 contig greater than 100 kb. Assembly statistics are summarized in Table 1. Based on CheckM, there is no contamination or heterogeneity in the genome. To further assess completeness, we generated a list of single-copy essential genes and screened them against the *C. lentocellum* DSM 5427 genome and against the draft genome, and the presence and absence was normalized the value to the *C. lentocellum* DSM 5427. Based on this more conservative method, it is more likely to be 92% complete. The genome encodes 54 tRNAs with at least one tRNA for all 20 amino acids (Table 2). We identified five rRNA operons containing 16S, 23S, and 5S rRNA

genes, however, one operon contained two 5S rRNA genes. There was one stand-alone 16S rRNA gene and a stand-alone 23S rRNA gene present in the genome, giving a total of 6 of each rRNA genes. The stand-alone 16S and 23S genes, as well as the rRNA operon with two 5S genes were present in every assembly that employed PacBio reads, making it unlikely that it represents a sequencing or assembly error. This unusual operon arrangement is possibly due to the high recombination rates found in type B (Arroyo, unpublished). No plasmids were detected in the genome, summarized in Figure 1.

Table 2. Count of tRNAs present in the '*Ca. Epulopiscium viviparous*' genome.

Amino acid	tRNA count	Amino acid	tRNA count
L-alanine	4	L-leucine	5
L-arginine	4	L-lysine	2
L-asparagine	3	L-methionine	5
L-aspartate	2	L-phenylalanine	1
L-cysteine	1	L-proline	2
L-glutamate	2	L-serine	4
L-glutamine	2	L-threonine	3
glycine	4	L-tryptophan	1
L-histidine	1	L-tyrosine	1
L-isoleucine	4	L-valine	3

Numbers are represent those identified by IMG.

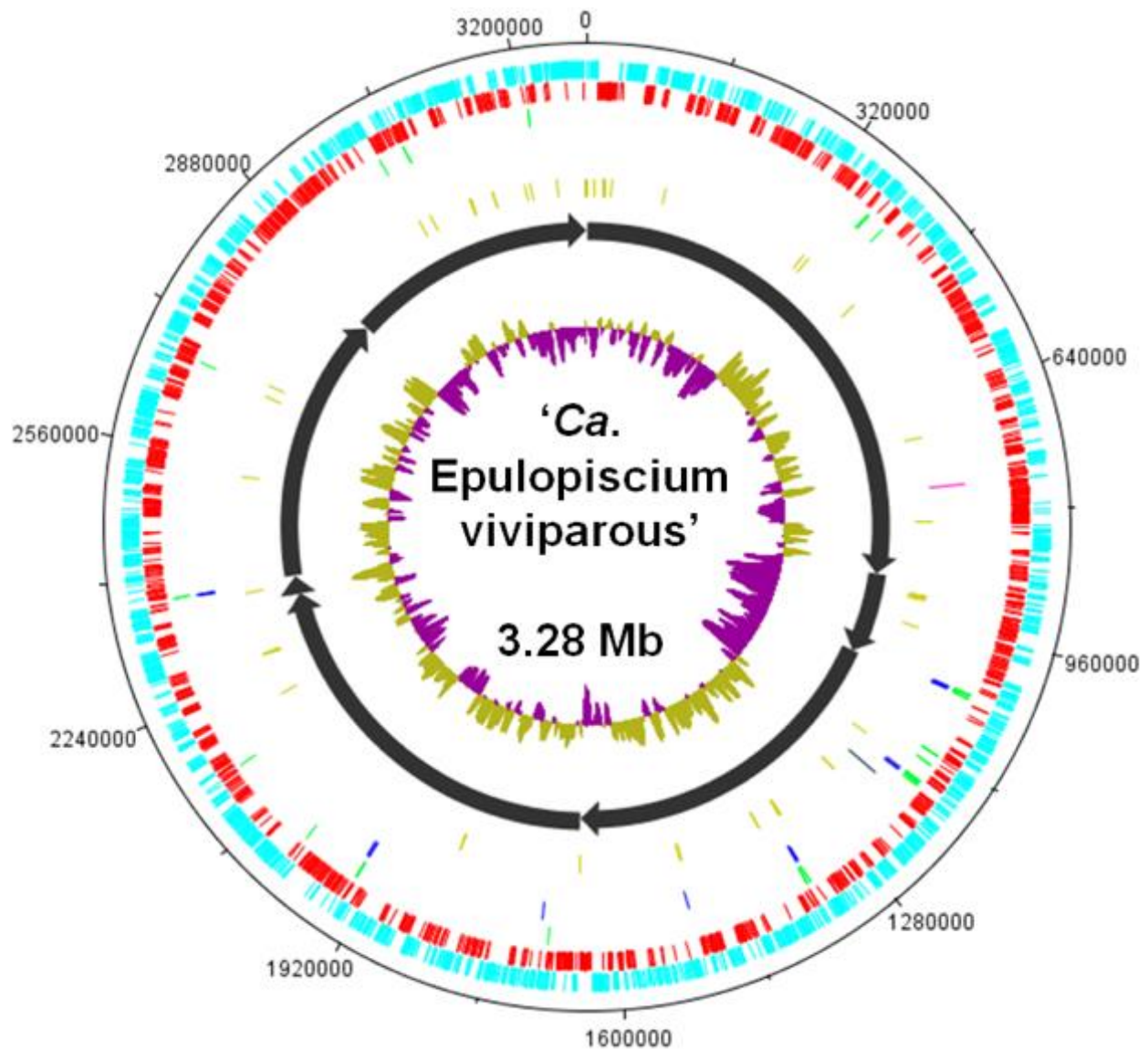


Figure 1. Circular view of the ‘Ca. *Epulopiscium viviparous*’ genome. Features of the 3.28 Mb genome are labeled in the circular representation of the genome. To generate this, the 7 contigs were concatenated, but unordered. From outer to inner; the light blue represents the coding DNA sequences of the continuous strand, and the red marks represent the coding DNA sequences on the lagging strand. Below this in green are the tRNAs, followed by the rRNAs in royal blue. In the next layer, *dnaA* (representative of the *oriC*) is represented by a dark green tick, while *xerC* (indicative of the terminus) is represented by the pink tick. The next circle composed of gold ticks depicts the repeat regions containing imperfect tandem repeats that occur more than 10 times, and the repeat units are at least 9 bp long. The repeated sequence lengths are all multiples of 3. The gray arrows depict each contig with the arrow head representing the end of the contig. The innermost circle represents the G+C skew of the contigs with the mustard color representing overabundance of G+C while the purple represent the lower abundance of G+C content.

The ‘*Ca. Epulopiscium viviparous*’ genome is circular

Due to the massive amount of DNA present in type B cells (28), the complex cytoarchitecture of the type B nucleoids (24), and the fragmentary nature of the previously published genome, the composition of the type B was unclear. A single circular chromosome is common in bacteria. Across the bacterial domain circular chromosomes replicate in a bidirectional manner from a single origin of replication (*oriC*). The *oriC* region of type B is conserved with *B. subtilis* and *C. lentocellum* DSM 5427, containing *dnaA*, *dnaN*, *yaaA*, *recF*, *gyrB*, and *gyrA*. *B. subtilis* has *yaaB* in between *recF* and *gyrB*, but this is missing in both *C. lentocellum* DSM5427 and type B. In the 5' direction from *dnaA*, all 3 genomes contain *rnpA*, *mnmE*, *mnmG*, *rsmG*, and *noc*, however, unlike in *B. subtilis*, both type B and *C. lentocellum* have a hypothetical protein and *yidC* directly after *rnpA*. *B. subtilis* and *C. lentocellum* contain *jag* in this cluster, but it is absent from type B. After ordering the contigs, we were able to observe a GC skew (58) indicative of circular bacterial genomes and identified a *xerC* homolog in proximity to the suspected terminus. For verification of this method, we used the closed *C. lentocellum* genome and were able to determine a terminus region in a similar manner, validating our approach. This finding further substantiates that the genome is circular. We found no evidence of a plasmid in the assembled genome.

The ‘*Ca. Epulopiscium viviparous*’ genome encodes for a form of reproduction derived from endosporulation

Type B reproduces from a derived form of endosporulation, in which two or more live intracellular offspring are produced in a diurnal cycle (27). The development stages exhibited in type B have hallmarks of the canonical sporulation pathway found in *B. subtilis* and other

endospore-forming Firmicutes (23, 32, 59). A comprehensive set of 147 core sporulation genes shared by *B. subtilis* and Clostridia was previously determined in our laboratory and used to interrogate an older and less complete *Epulopiscium* sp. type B draft genome (27). In the previous study, 57 putative sporulation homologs were found, with at least 5 genes found for each of the 4 sporulation specific σ -factor and Spo0A regulons (27). We were able to locate these 57 genes in our current assembly, as well as 6 more highly conserved sporulation genes. In the previous assembly, *spoIIP* was found on a pseudomolecule and was missing its 5' end, however, in the current assembly it is intact and next to *gpr* as was discovered by PCR (27). Further, we were able to locate putative homologs to gerW (formerly *ytfJ*) of the σ^F regulon (60), *yaaH* (61) and *yngI* of the σ^E regulon (62), and *ytcC* (63), *ytIC*, and *ytID* of the σ^K regulon (64), giving the genome 63 sporulation specific protein homologs. Unlike *B. subtilis*, Clostridia do not use the phosphorelay system to activate Spo0A, the master regulator of the initiation of sporulation. Rather Clostridia use sensory orphan histidine kinases (HKs) to directly phosphorylate Spo0A (65, 66), and different clostridia have multiple HKs that respond to environmental conditions to initiate sporulation. We were able to locate 2 orphan HKs in the genome, one of which is similar to a HK that activates Spo0A in *Clostridium acetobutylicum*, *C. thermocellum*, *C. botulinum*, and *Peptoclostridium difficile*, suggesting that it is playing a similar role in activating Spo0A in type B. The presence of sporulation genes in our high quality draft genome from a different type B population than what was used in a previous study (27) reaffirms the finding that intracellular offspring formation is a derived form of endosporulation, and presents a candidate sensory HK for the activation of this process.

The '*Ca. Epulopiscium viviparous*' genome encodes for multiple giant genes

It has been suggested that due to the cost of producing genes (67), large genes (>5 kb) are rare in bacteria and archaea. However, Reva and Tümmler demonstrated that when examining the ORFs of fully sequenced genomes, large genes do occur at a frequency of 0.2%. This frequency is higher in '*Ca. Epulopiscium viviparous*' as 0.8% of the genes are greater than 5 kb. Of these 21 large genes, 3 are longer than 15,000 kb, and 6 are longer than 10 kb. The largest gene is 20,610 bp, encoding for a protein 6,870 amino acids in length. The protein is secreted, contains transmembrane domains, as well as a fibronectin III domain, a vinculin domain, and a parallel β -helix domain found in pectin and rhamnogalacturon lyases. The fibronectin III and vinculin domains may play a role in adhering to host epithelial cells. The other large, secreted genes over 9 kb in length lack any discernable functional domains, but do appear to contain transmembrane helices. They have amino acid signatures reminiscent of giant surface proteins in other bacteria, in that they are almost completely lacking cysteine residues, are low in arginine, but are enriched for alanine, threonine, aspartate, and glutamate (68). Like other giant genes, some of the genes are highly repetitive, with imperfect direct repeats, or repeated sequences concentrated in a certain part of the gene (68).

The role of these giant surface proteins is unclear, but they may play roles in defense, particularly for evading viruses (69, 70). This may be occurring as we were unable to locate any prophage within the genome. In *Synechococcus* WH8102, the giant surface protein SwmB reduces predation by a predatory dinoflagellate (71). Type B is preyed upon by the ciliate *Balantidium jocularum* in the gut of *N. tonganus* (72), and some of these surface proteins may serve an analogous function to SwmB, helping type B evade predation. Giant surface proteins in the green-sulfur bacteria *Chlorobium chlorochromati* CaD, which forms a phototrophic consortia with

chemotrophic β -proteobacteria, potentially serve as symbiosis factors, allowing for the smaller *Chl. chlorochromati* to aggregate on the larger cells as epibionts (73). Due to type B's large size, these large surface proteins may allow for the attachment of smaller bacteria to the surface as well.

Four of the 6 genes that are over 10 kb in length do not have apparent signal peptide sequences or transmembrane domains, suggesting that these proteins are cytoplasmic. InterproScan identified immunoglobulin-like (Ig) fold domains in 3 of the 4 putative proteins. The 2nd largest gene in the genome is 20,502 bp, and encodes a non-secreted 6,833 amino acid protein with 31 Ig-fold domains. The 3rd largest gene in the genome is 16,596 bp, encoding for a 5,532 amino acid protein with 21 Ig-fold domains. The repetition of Ig-like domains is reminiscent of the giant adhesin SiiE in *Salmonella enterica* (74). We speculate that these non-secreted giant genes may play a unique role in type B, where they potentially function as adhesins that anchor daughter cells to confined locations in the cytoplasm during development. Since type B can form up to 12 daughter cells, not all daughter cells are formed at the poles. Using these proteins as anchors could be a way the mother cell delineates where laterally positioned daughter cells develop within the cytoplasm.

'*Ca. Epulopiscium viviparous*' encodes for the general secretory and the ESX secretion system

Like other Gram-positive bacteria, type B relies on the general secretory pathway for exporting proteins. Proteins are targeted for export as they possess a hydrophobic N-terminal signal peptide, which is recognized by the signal recognition particle (SRP) (75)., Type B encodes *ffh* and *hbsU*, which are part of the SRP, and *ftsY* which is the SRP receptor. The genome further encodes for the SecAYEG translocon, as well as the accessory factors SecD and SecF. Type B

only encodes for one YidC protein and like most Gram-positive bacteria, does not produce a SecB homolog or CsaA, which may serve a similar chaperone function to SecB in *B. subtilis* (75). The genome encodes for 3 putative signal peptidases for the cleavage of the signal peptide, after modification by Lgt, and the LspA lipoprotein signal peptidase, both of which are present in the genome (75). The type B genome also codes for RasP, which is used to degrade the signal peptides. There is no twin-arginine-translocation system present in the genome, suggesting that the complete general secretory pathway is the major protein translocation system.

The ESX or Type VII secretion system is found in Gram-positive bacteria, and its function is specialized for the species in which it is located. In some species it functions in virulence (76) or spore formation (77), but in other species like *B. subtilis* its function is unknown. The *B. subtilis* system is characterized by having an FtsK/SpoIIIE family ATPase YukBA, which is encoded in the *yukEDCBA* operon containing structural genes and the secretion target, the small peptide YukE, (78). *B. subtilis* also contains the *yfiA* ESX associated gene in another location on the chromosome (79). We were able to uncover a modified form of this type VII secretion system operon in type B, which contains *yukBA*, *yukD*, and 5 *yukE* homologs, suggesting type B has a variation of the secretion system. We were unable to find *yueBC* homologs or *yukC*, which are all found in *B. subtilis* (78), but did find a putative *yfiA* homolog. It is unclear why '*Ca. Epulopiscium viviparous*' encodes for this secretory system and why it has multiple copies of the secreted protein.

'*Ca. Epulopiscium viviparous*' is flagellated and chemotactic

Type B cells contain peritrichous flagella and are highly motile. As they swim they rotate along their long axis and move in a straight line. When they reverse their direction the rotation of the cell also reverses (24). This motility likely allows for the cells to maintain position in the gut,

but may also allow for cells to torpedo into chunks of algal fragments embedding themselves into the algae to better access and degrade the present polysaccharides. Stationary cells often rotate their flagella, likely for the purpose of refreshing their immediate environment (24).

Bacterial flagella are complex molecular machines composed of a basal body, which anchors the flagella into the membrane, a hook, which acts as a joint to generate torque, a filament, which functions as a propeller, and a stator that translates proton or ion motive force into torque (80-83). For assembly to occur multiple components need to be secreted through a flagellar specific type III secretion system (fT3SS) as well. In *B. subtilis*, most of the genes necessary for flagellar formation are located in a large 27kb operon containing 31 genes (Figure 2) (81). This *fla/che* operon is highly conserved in type B (Figure 2), as it contains all the homologs present in *B. subtilis*, allowing for type B to assemble the basal body, the fT3SS, and hook. Distinct from *B. subtilis*, the type B operon contains *motA* and *motB* in this operon. *B. subtilis* has 2 stators, MotAB, which generates torque through PMF and MotPS, which generates torque through SMF (84), however, type B only has one set, suggesting it uses either PMF or SMF as energy. The only other divergence in this operon is that the type B operon lacks an *swrB* homolog. Like *B. subtilis*, the genome has *flgMNKL*, *fliW*, and *csrA*, genes involved in biosynthesis and regulation, all in an operon, but the type B operon contains 2 *flgK* proteins and does not have *yviE* found in *B. subtilis* (81). Unique to type B is that the operon also contains a gene encoding for a FliB family protein. We were unable to detect *flhO* and *flhP* homologs, of the *B. subtilis* rod structural genes, but did detect *flgF* and *flgG* homologs in a gene cluster, which are also rod structural genes. Type B also encodes for the *fliD* cap and *fliS* chaperone in cluster together, and has another copy of *fliS*, located in a different location in the genome.

The filament is composed of thousands of subunits of the protein flagellin. We identified 9 putative flagellin genes, with 7 located adjacent to one another in an operon next to the *fla/che* operon. The other two flagellin genes are located in separate operons, with one adjacent to a *fliS*. The presence of 9 different flagellin proteins suggests a potential for them to play multiple roles for type B. The large variety of possible flagellin expression patterns may be important for evading detection by the host immune system, allowing for its maintenance and survival as a gut symbiont (85). It may also be that some flagellins are better suited for different gut conditions, and the types of flagellin used in assembly are environmentally controlled (85). The variability in epitopes type B can create may also allow for it to adhere to a wider array of host derived molecules or other gut contents (86). The high avidity generated with multiple diverse repeating epitopes (86) could provide better binding strength which may be necessary for stable adherence of such a large cell to host tissue and other compounds in the gut. Taken together, the type B has the ability to produce multiple flagella varying in their flagellin content, which may be important for gut maintenance and niche exploitation.

Type B appears to be chemotactic, allowing it to have guided motility towards chemo-attractants and away from chemo-repellents. The *fla/che* operon contains the chemotaxis genes *cheA*, *cheB*, *cheC*, *cheD*, *cheW*, and *cheY* (Figure 2). The genome includes 2 copies of the methyltransferase *cheR*, and *cheV*. These genes provide type B with the ability to employ the 3 chemotaxis adaptation systems exhibited in *B. subtilis* (87). Integral to chemotaxis are methyl-accepting chemotaxis proteins (MCPs), which are chemoreceptors that transduce the signals directing motility. Typically, MCPs have a ligand-binding domain (LBD), which can adopt a variety of folds, one or multiple transmembrane helices, a HAMP region, and a cytoplasmic signaling domain (88). We were able to identify 3 canonical MCPs, 2 of which have calcium

channels and chemotaxis receptors (Cache) domains in their LBD, with the other having a taxis to aspartate and repellents (Tar), taxis to serine and repellents (Tsr) superfamily domain in its LBD. We identified 7 other putative MCPs. The genomic content suggests that ‘*Ca. Epulopiscium viviparous*’ is chemotactic to a variety of compounds in the gut of *N. tonganus*.

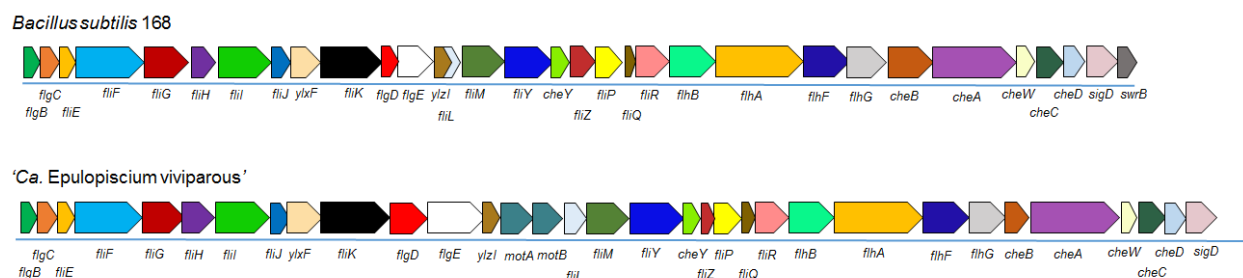


Figure 2. Comparison of the *fla/che* operon between ‘*Ca. Epulopiscium viviparous*’ and *B. subtilis* 168. The *B. subtilis* 168 operon is on top while the type B operon is on the bottom. Each homologous gene pair shares the same color when comparing between both species.

‘*Ca. Epulopiscium viviparous*’ can produce a cell wall containing meso-diaminopimelate and may produce cell wall-associated glycopolymers

Typically, Gram-positive bacteria produce cell walls containing L-lysine in their peptidoglycan (PG) side chains, however, some Firmicutes synthesize a cell wall that contains the lysine biosynthesis intermediate meso-diaminopimelate (DAP) instead (89). The type B genome encodes for UDP-*N*-acetylmuramoylalanyl-D-glutamate 2,6-DAP ligase for the incorporation of meso-DAP during the synthesis of lipid II, as well as the repertoire of genes to make precursors like N-acetylglucosamine and N-acetylmuramic acid, as well as all the genes involved in lipid II synthesis. Lipid II is the major precursor molecule for PG and is synthesized in the cytoplasm, then flipped out of the membrane most likely by the flippase MurJ or FtsW (90, 91). We identified putative homologs of both *murJ* and *ftsW* in the genome. We identified a penicillin-binding-protein (PBP) family 1a transglycosylase for the synthesis of PG chains. To accomplish crosslinking between the D-alanine and DAP of 2 different chains, type B encodes for 2 putative class B PBPs,

one a *pbpA* homolog in a gene cluster next to a gene encoding an FtsW-RodA-SpoVE family flippase, and the other a PBP2B transpeptidase in an operon with *mreB*, *mreC*, *mreD*, *minC*, *minD*, *minE*, and gene encoding an FtsW-RodA-SpoVE family flippase. The genome also includes a low molecular weight D,D-carboxypeptidase PBP to potentially help control the degree of cross-linking in the PG (92). Further 11 putative β -lactamases are present in the genome.

Type B likely produces cell wall-associated glycopolymers. We were able to identify some of the genes conserved between *Bacillus* and *Staphylococcus* spp. involved in the biosynthesis of teichoic acids. Three genes identified in the type B genome include *tagO*, *tarA/tagA*, and *mnaA* which are used to form the N-acetylmannosamine- β -(1,4)-N-acetylglucosamine linkage unit,(93). However, the conserved *tagB*, which transfers glycerol phosphate to the linkage unit was not identified in type B. We were unable to locate most of the other known *tar* and *tag* genes found in *Bacillus* and *Staphylococcus* spp., but did uncover an operon that likely functions in cell wall-associated glycopolymer synthesis. The operon contains a putative homolog of *tagT* for the potential transfer of a lipid linked glycopolymer to peptidoglycan. The operon also contains a phosphodiester glycosidase (NAGPA family protein), 2 glycosyltransferase (GT) family 4 proteins, and a *murJ* flippase homolog, which may function in translocating of the glycopolymer across the membrane. Within the operon there are also 2 membrane associated hypothetical proteins as well as a GT9 protein (which performs a glycosyltransferase reaction in LPS biosynthesis in Gram-negative bacteria), a gene annotated as an O-antigen polymerase, and the *tarA* homolog. As type B does not produce LPS, the 2 genes annotated to be associated with LPS biosynthesis are likely performing similar functions in the synthesis of the type B extracellular glycopolymer. The genome also encodes for a GT8 protein which may be involved in glycopolymer synthesis as well.

We were able to detect a few putative genes involved in the synthesis of lipoteichoic acids, but not a complete set. We detected a putative diacylglycerol kinase, as well as a GT4 with homology to *lafA*, but were unable to detect *lafB*. LafA and LafB function together to produce the glycolipid anchor of teichoic acids in *Listeria monocytogenes* (94), making it unclear if type B can make the anchor. *B. subtilis* only uses one protein, YpfP, to generate the glycolipid anchor (95), but no homologs were present. It is possible the type B LafA homolog alone may be suitable as the anchor. We were able to detect *pgsA* and *cdsA* homologs, which play a role in lipid recycling in lipoteichoic acid synthesis. Lipoteichoic acid synthesis is not well understood in Clostridia, and there may be unidentified genes present involved in this process. Taken together it is unclear the exact nature of the glycopolymers present on the outside of type B cells, but it does appear that the genome encodes the ability to make complex extracellular structures.

Poly- γ -glutamate, composed of D- or L-glutamate, or a combination of the enantiomers is produced by some Firmicutes and is either secreted or anchored to the cell wall, forming a capsule. Secreted poly- γ -glutamate allows for adaptation to adverse environments, or serves as a nutrient during starvation (96). *B. anthracis* produces a capsule which serves as a virulence factor, and requires the *capBCADE* operon for this (96). We detected a homolog of *capA* in the type B genome, in an operon with 3 other genes of unknown function. Due to the presence of this capsule biosynthesis gene, it is possible type B produces a form of a poly- γ -glutamate capsule, or once had the ability to make a capsule.

Figure 3. The metabolic and biosynthetic landscape of ‘*Ca. Epulopiscium viviparous*’. The image depicts the metabolic and biosynthetic potential of a ‘*Ca. Epulopiscium viviparous*’ cell. The orange inner line circling the cell depicts the membrane, while the thick gray line surrounding the cell depicts the PG layer. Specific metabolic pathways are highlighted in colored boxes with key intermediates present. Degradative steps are depicted by black lines, and dotted lines stand for missing steps of pathways. Carbohydrates grouped together assimilate into the same intermediate, but not necessarily through the same pathways (reference the text for more detail). In sodium motive force generation box, the membrane bound oxaloacetate decarboxylase is labeled Oaa decarb, and the gray arrow from Oaa to Oaa represents 2 sources of the compound. Above the cell are cartoon depictions representative of the host algal diet; to the left a filamentous brown algae, the center, a red algae, and to the right, a *Caulerpa* green algae. Polysaccharide degrading enzymes are depicted as open circle. The brown one depicts alginat lyases and laminarinases. The red depicts the variation of agarases, porphyranases, and carrageenases. The green depicts the xyloglucanases and pectate lyases. The red-green open circle on both the red and green algae depict the β -mannanases, β -galactanases, and GH42’s active on these compounds. The oligosaccharides released feed into the red box, where they are presumably imported (designated by the 3 brown importers), and degraded by the cytoplasmic enzymes. Sugar transporters are colored in brown. The narrow diamonds depict PTS systems, the fatter diamonds depict multiple sugar transporters, and the hexagonal transporters depict specific transporters for specific hexoses. Blue squares depict amino acid importers and exchangers, with amino acids coded by their single letter abbreviations in blue text. Blue arrows depict biosynthesis pathways of the amino acids, and light blue rectangles represent transporters for urea and NH_3 . Red octagons are vitamin transporters, with the specific B-vitamins labeled in red text, and red arrows depict their biosynthesis. Purple arrows illustrate biosynthesis of NTPs and dNTPs. Inorganic compounds are in green text, and with the green transporters being ion translocators. P’s next to carbohydrates stand for phosphate. D-rib is D-ribose, D-rbl is D-ribulose, D-sep is D-sedoheptulose, D-ery is D-erythrose, G3P is D-glyceraldehyde-3-phosphate, D-fru is D-fructose, gal is galactose and 3,6-an depicts the anhydro versions of galactose. D-mtl is D-mannitol, D-gat is D-galactitol, D-xyl is D-xylose, L-ara is L-arabinose, D-glu is D-glucose, suc is sucrose, mal is maltose, and tre is trehalose. D-man is D-mannose, mng is digeneaside, D-glc is D-gluconate, L-fuc is L-fucose, cit is citrate, Oaa is oxaloacetate, BCAAs is branched-chain amino acids, M-pre is methionine precursors, and PolyP is polyphosphate.

‘*Ca. Epulopiscium viviparous*’ encodes for a complete EMP and reductive Pentose Phosphate Pathway, but incomplete TCA cycle

Like other members of the Clostridia, type B appears to rely on fermentation for energy conservation (97), as it lacks the genomic components for respiratory complexes. The central metabolism of ‘*Ca. Epulopiscium viviparous*’ is summarized in Figure 3. The genome encodes a complete Embden-Meyerhof-Parnas (EMP) glycolytic pathway. In the EMP pathway, type B

codes for an ATP-dependent phosphofructokinases, as well as the pyrophosphate-dependent diphosphate-fructose-6-phosphate 1-phosphotransferase, allowing it to generate fructose-1,6-bisphosphate using 2 different phosphate donors. The pyruvate generated from the EMP pathway can be fermented to generate ATP and NAD^+ . Two pyruvate formate lyases are present in the genome to convert pyruvate to acetyl-CoA. Phosphate acetyltransferase and acetate kinase are located next to each other in the genome to convert acetyl-CoA to acetylphosphate and then acetate, generating ATP. The genome contains all of the genes necessary for fermentation of acetyl-CoA to acetaldehyde and then ethanol by acetaldehyde dehydrogenase, regenerating NAD^+ . Multiple alcohol dehydrogenases are also present in genome. Methylglyoxal is a toxic product of glycolysis, and the genome encodes for a methylglyoxal reductase to convert it to acetol, which can then be converted to L-1,2 propanediol, with NAD(P)^+ generated in each step. The genome also encodes for the complete set of genes for the Non-oxidative branch of the pentose phosphate pathway (PPP).

Like *C. lentocellum* DSM 5427 and other Clostridia (98), the type B genome encodes an incomplete TCA cycle and is missing the genes to convert α -ketoglutarate to fumarate, but it has the typical genes to convert oxaloacetate to α -ketoglutarate (Figure 3). Despite the cycle being truncated, it is still able to produce a necessary intermediate for amino acid metabolism, and it likely functions to link carbohydrate metabolism to amino acid biosynthesis, but serves no function for energy conservation.

‘*Ca. Epulopiscium viviparous*’ can generate a sodium motive force through oxaloacetate decarboxylase

Oxaloacetate is an important molecule in bacteria as it serves as a precursor for L-aspartate synthesis. Type B has the potential to generate this molecule and ATP through its production from

phosphoenol pyruvate (PEP) by PEP carboxykinase (99). Oxaloacetate can be further decarboxylated to pyruvate using the Na^+ dependent, membrane bound oxaloacetate decarboxylase, which is encoded for by the genome. With this enzyme, during the decarboxylation, Na^+ is pumped out of the cell generating a Na^+ motive force (SMF) (100). The enzyme is generally composed of 3-4 subunits, and type B encodes for it in an operon containing 4 genes for the subunits, as well as an acetyl-CoA carboxyltransferase. Oxaloacetate decarboxylase is used by some Enterobacteriaceae for anaerobic growth on carboxylates as the sole carbon source (101, 102). *Klebsiella pneumoniae* is able to ferment citrate to oxaloacetate for its subsequent decarboxylation (102, 103). This is achieved through the citrate lyase complex, which cleaves it to oxaloacetate and acetate. This reaction also requires the production of a prosthetic group via CitX and citrate lyase ligase, and these genes are all located in an operon in '*Ca. Epulopiscium viviparous*' (102, 103). The citrate lyase operon includes a putative tricarboxylate transporter for the potential uptake of citrate, as well as aconitase and 3-methylitaconate isomerase. This provides type B with another source of oxaloacetate for SMF generation. In its oxaloacetate decarboxylase operon, *K. pneumoniae* has the genes for a citrate sensing two-component system and a citrate transporter (102). For growth on citrate, *B. subtilis* has a citrate sensing two-component system along with the Mg^{2+} -citrate transporter CitM (104). CitM is present in the type B genome, however, we were unable to identify a citrate specific two-component system. This could be due to the fact that the other organisms are facultative aerobes and not obligate anaerobic fermenters like type B, which does not need to rely on external signals to switch metabolism to anaerobic fermentation. The pyruvate generated from the fermentation of oxaloacetate can be further fermented to acetate for additional ATP.

'*Ca. Epulopiscium viviparous*' encodes for a Na^+ -dependent F_1F_0 ATP synthase

Some anaerobes can use the SMF generated from their decarboxylases to synthesize ATP (105, 106). *Propionigenium modestum* generates a SMF from succinate fermentation (106) *Fusobacterium nucleatum* generates it from glutamate fermentation (107), and both species encoding for a sodium specific F₁F₀ ATP synthase. The F₀ intrinsic membrane domain is composed of an oligomeric ring of c subunits, and these subunits define the specificity for the translocated ion (107, 108). The Na⁺ translocating F₁F₀ ATP synthases of *F. nucleatum* and *Ilyobacter tartaricus* have conserved residues in their 89 amino acid c subunit, Gln/Glu³², Glu⁶⁵, Ser/Thr⁶⁶, and Tyr⁷⁰ that are necessary for Na⁺ binding (Figure 4A) (108). These residues are found in the sequences of the c subunits of other Na⁺ translocating bacteria (107, 108). The type B genome encodes for an F₁F₀ ATP synthase, with the genes for the ABC subunits of the F₀ domain and the αβγδει subunits of the F₁ domain all located in an operon. The putative type B c subunit homolog appears to be specific for Na⁺, as it has the conserved residues Glu³¹, Glu⁶⁴, Thr⁶⁵, and Tyr⁶⁹ (Figure 4A). The c subunit appears to be triplicated in the genome, as next to *atpE*, the gene for the c subunit, there is another copy of the gene that encodes for a protein with two AtpE domains. The first AtpE domain in this c subunit protein maintains the same important Glu, Glu, and Thr residues important for Na⁺ binding, but has a Phe instead of a Tyr residue (Figure 4B). The Tyr residue, along with Ser/Thr donate H bond to the Glu to keep it deprotonated, but it also likely relocates to allow for loading and unloading of the binding site (108). The Phe cannot donate the H-bond, but due to its similarity in structure it could likely play the same function of allowing for loading and unloading of the binding site. The second AtpE domain appears to be truncated, but maintains the first Glu residue, however, the residue analogous to Glu⁶⁴ is a Gln. This is the only discrepancy as it retains the important Thr and Tyr residues. It is unclear why there are 3 c subunits encoded for,

but it may be for structural purposes. The presence of this type of F₁F₀ ATP synthase makes it likely that type B produces some ATP using a SMF generated.

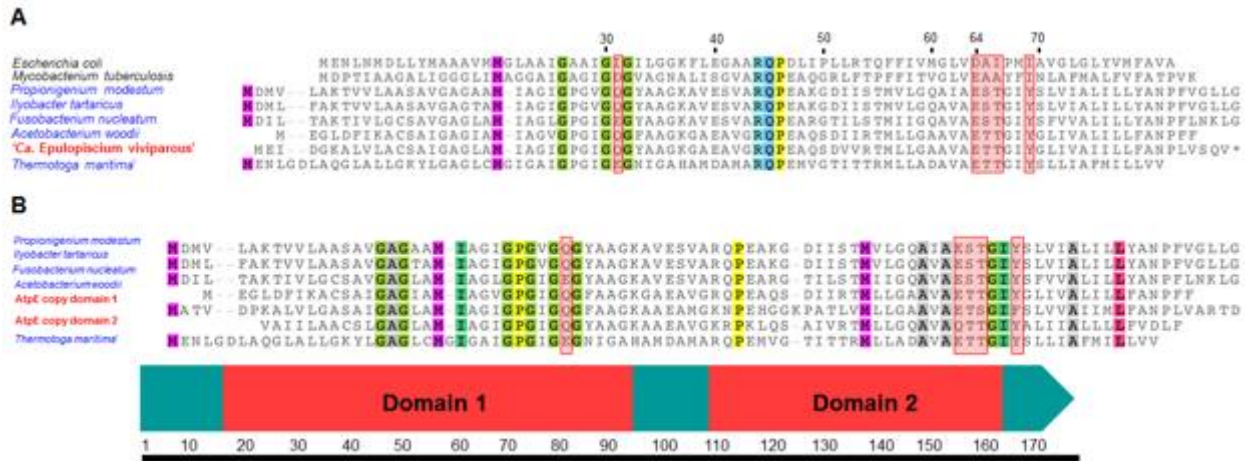


Figure 4. Analysis of the AtpE amino acid sequences of ‘*Ca. Epulopiscium viviparous*’. (A) depicts the alignment of the amino acid sequence of ‘*Ca. Epulopiscium viviparous*’ AtpE with species that have an H⁺ dependent (species in black text) or an Na⁺-dependent AtpE (species in blue text). The numbering scheme is based on the type B sequence. Important residues necessary for Na⁺ binding are boxed in red. H⁺ dependent AtpE lack an E or Q at the 31 position and lack an S/T at the 65 position and a Y at the 70 position. (B) is an alignment of the amino acid sequence of the second copy of the AtpE found in the genome to the Na⁺ dependent species. This gene was broken up into 2 sequences as it contains 2 AtpE domains (domain 1 and domain 2). The same important residues in (A) are highlighted in red boxes in (B). In the position analogous to the 64 position, the 2nd domain has a Q instead of an E. In the position analogous to the 69 position, 1st domain has an F instead of a Y. Underneath the alignment is a schematic of the 179 amino acid protein, with both domains signified.

The ‘*Ca. Epulopiscium viviparous*’ genome is enriched for carbohydrate metabolism

We performed a COG analysis of the type B assembly (Table 3) and compared those results to analyses of genomes of other members of the Lachnospiraceae. Those selected for the analysis included 3 free-living species, as well as one human gut isolate and one rumen associated member of the Lachnospiraceae XIVb cluster. Also included are 3 human gut associated members of the Lachnospiraceae XIVa cluster. This selection allowed for a comparison between both free-living and gut-microbiota to see if there were any distinctions based on niche (Figure 5). Compared to the other Lachnospiraceae genomes selected, type B devotes the highest proportion of its genome

to carbohydrate transport and metabolism, as 15.43% of all proteins with COGs have functional carbohydrate metabolism domains (Figure 5). Every other genome had $\leq 13\%$ devoted to this COG, with most closer to 8%. The genomes whose carbohydrate COG proportions were more than 8% tended to be gut-associated (except *Cellulosilyticum* sp. I15G10I2, which is free-living. This suggests that the enrichment for carbohydrate metabolism in type B is unique to its niche as a gut symbiont of *N. tonganus* and not just a general product of being host associated. Comparatively, type B also had the highest percentage of COGs associated with amino acid transport and metabolism, ion transport, coenzyme transport and metabolism, and general function (Figure 5).

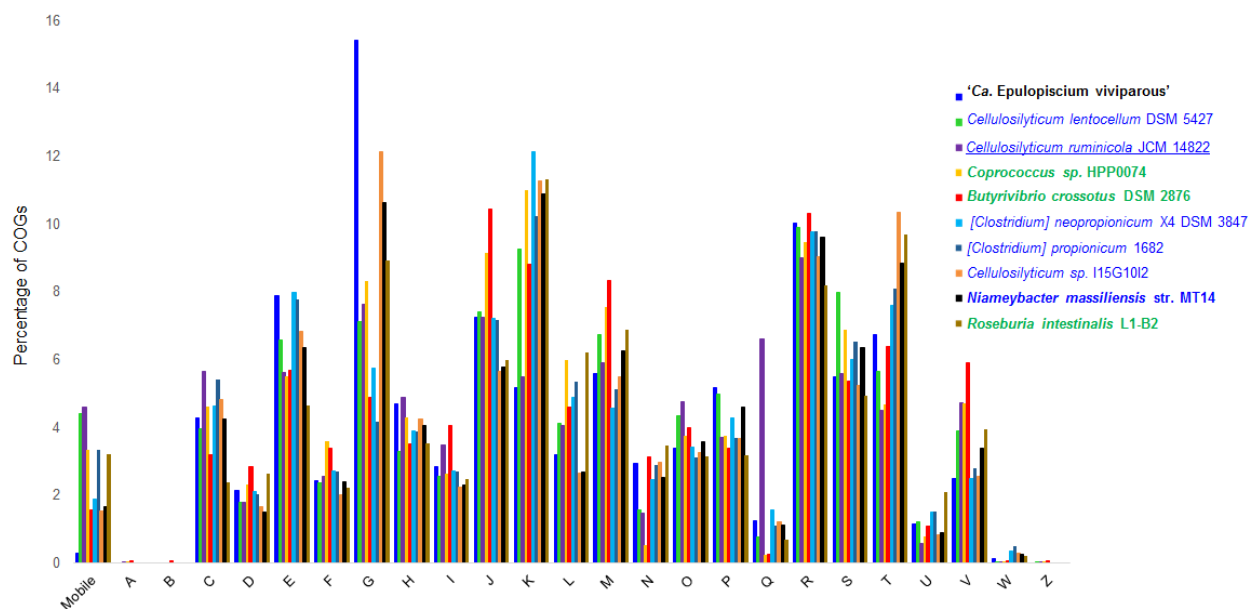


Figure 5. COG comparison between ‘*Ca. Epulopiscium viviparous*’ and its relatives. COG analysis reveals that ‘*Ca. Epulopiscium viviparous*’ is enriched for carbohydrate metabolism. Members of the Lachnospiraceae XIVa cluster are in green text, and members of the Lachnospiraceae XIVb cluster are in blue text. Human gut-associated species are in bold font, and the rumen associated species is underlined. Refer to Table 2 for COG category codes.

Table 3. Cluster of Orthologous Genes profile of '*Ca. Epulopiscium viviparous*'

COGcluster	Code	Percentage of COGs
Mobilome: prophages, transposons	Mobile	0.279
RNA processing and modification	A	0
Chromatin structure and dynamics	B	0
Energy production and conversion	C	4.277
Cell cycle control, cell division, chromosome partitioning	D	2.139
Amino acid transport and metabolism	E	7.903
Nucleotide transport and metabolism	F	2.417
Carbohydrate transport and metabolism	G	15.435
Coenzyme transport and metabolism	H	4.695
Lipid transport and metabolism	I	2.836
Translation, ribosomal structure and biogenesis	J	7.252
Transcription	K	5.160
Replication, recombination and repair	L	3.208
Cell wall/membrane/envelope biogenesis	M	5.579
Cell motility	N	2.929
Posttranslational modification, protein turnover, chaperones	O	3.394
Inorganic ion transport and metabolism	P	5.160
Secondary metabolites biosynthesis, transport and catabolism	Q	1.255
General function prediction only	R	10.042
Function unknown	S	5.486
Signal transduction mechanisms	T	6.741
Intracellular trafficking, secretion, and vesicular transport	U	1.162
Defense mechanisms	V	2.510
Extracellular structures	W	0.139
Cytoskeleton	Z	0

Due to its diverse carbohydrate metabolism capacity, we conducted a Carbohydrate Active enZymes (CAZy) analysis of the type B genome. The genome contains 131 glycoside hydrolase (GH) CAZymes from 38 different families, 8 polysaccharide lyases (PL) from 2 families, and 26 carbohydrate esterases from 9 different families (Table 4). The degree and diversity of CAZymes present in the genome is slightly less than what is observed in *Zobellia galactivorans* Dsij, a heterotrophic marine bacterium isolated from red algae with the most known CAZymes of any

known heterotrophic marine bacterium (109), but more than what is observed in other free living algal associated marine heterotrophic bacteria such as *Formosa agariphila* KMM 3901^T (109, 110). The abundance of COGs and the diversity of CAZymes associated with carbohydrate metabolism, and the comparable presence of these enzymes to free-living algal degrading bacteria, further support the long-standing hypothesis that type B functions as a gut symbiont that helps degrade recalcitrant polysaccharides (27) to produce short chain fatty acids for its host. Table 5 summarizes the polysaccharide degradation abilities of 'Ca. Epulopiscium viviparous' and Table 6 summarizes its sugar assimilation capabilities.

Table 4. CAZYme profile of ‘*Ca. Epulopiscium viviparous*’

CAZYme family	No. in the genome	CAZYme family	No. in the genome
GH2	18	GH91	1
GH3	3	GH109	10
GH4	4	GH110	2
GH5	2	GH114	1
GH16	10	GH117	6
GH18	2	GH120	1
GH23	1	GH125	1
GH26	3	GH127	1
GH28	2	GH129	1
GH29	6	GH13	7
GH30	2	GH130	1
GH31	2	GH133	1
GH32	2	PL1	1
GH35	1	PL9	7
GH36	2	CE1	6
GH37	1	CE2	1
GH38	2	CE4	5
GH42	2	CE6	1
GH43	3	CE7	1
GH50	5	CE8	3
GH53	1	CE9	6
GH65	1	CE10	2
GH74	15	CE14	1
GH77	1	Total GHs	132
GH78	1	Total PLs	8
GH86	7	Total Ces	26

‘*Ca. Epulopiscium viviparous*’ has the ability to degrade polygalactans found in the cell walls of red algae

Naso tonganus has a diet dominated in both thallate and filamentous red algae (18, 111) agar and carrageenans are a major constituent of their cell walls (112-115). Agars are galactans with backbones consisting of alternating 3-O-linked β -D-galactose and a 4-O-linked α -L-galactose, with the L monomer being 3,6-anhydro- α -L-galactose in agaroses and α -L-galactose-6-sulfate in porphyrans (112). They vary from species in the amount and class of modifications

present on the monomers (115). The type B genome encodes for 12 putative β -agarases to degrade agarose by cleaving the β -(1,4) linkage (112). Agarases are typically found in multiple GH families including GH16, GH50, and GH86 (112, 116). GH16 and GH86 family agarases are usually endo-agarases that produce neoagarotetraose and neoagarohexaose oligomers, with some GH86 agarases producing neoagarobiose (116). GH50 agarases are typically exo-agarases that generate neoagarotetraose and neoagarobiose (112, 117). Of the 12 putative agarases, 2 were GH16, 5 were GH50, and 5 were GH86 agarases, and they tended to be found in clusters of 2 or 3 based on GH family. The GH16 agarases, which are located in the same operon, have carbohydrate binding module 6 domains (CBM6), which allow the enzymes to attach to algal cell walls (112). Also included in this operon is an exported GH42 β -galactosidase, which may be further degrading the oligomers generated from the agarases. The neoagarobiose is degraded by a group of GH117 α -neoagarobiose hydrolases (α -NAH) to generate β -D-galactose and 3,6-anhydro- α -L-galactose (112, 116). All six of these enzymes are cytosolic, like that of *Saccharophagus degradans* 2-40 (118), meaning that the neoagarobiose needs to be imported into the cell, which occurs via an unknown transport mechanism (112). One of the α -NAH is located in an operon that includes the genes involved in the degradation of 3,6-anhydro- α -L-galactose into 2-keto-3-deoxygluconate 6-phosphate (KDPG) in other bacteria (119), while another is located next to 2 sulfatases, possibly for the removal of sulfur groups found in porphyran. The diversity of agarases present in the genome may be due to different red algae modifications type B cells encounter in the guts of their host. Certain agarases may be able to accommodate certain modifications that others cannot.

We identified 4 GH16 proteins clustered together in the genome, with two likely being putative β -porphyranases, which also targets the β -(1,4) linkage of porphyran (112). The other two GH16 proteins appear to be cytosolic, and may be involved in the degradation of

neoporphyrbiose. We furthered identified an operon containing two putative GH86 endo- β -porphyranases (117), along with a putative cytoplasmic GH43 hydrolase, a non-secreted GH91 hydrolase, and an arylsulfatase. α -Neoporphyrbiose hydrolases and sulfatases specific for removing the sulfur from α -L-galactose-6-sulfate have yet to be discovered in agarolytic bacteria (112), however, it is possible that the arylsulfatase present in the operon performs the desulfatase function, to generate L-galactose. The L-galactose is likely metabolized as in *Bacteroides vulgatus*, as the putative pathway to convert L-galactose to D-tagaturonate is present in the genome (120). There are also 4 intracellular α -(1,3) galactosidases encoded for, and one of them may function to degrade the neoporphyrbiose to the monomers.

Carrageenans are also sulfated galactans however, they differ from agars in that both monomers are in the D conformation. Carrageenans are characterized by the amount and positions of sulfur esters present on both galactose units (114). We identified an operon containing a CUT1 family ABC transporter with two substrate binding proteins, a putative GH110 α -1,3-galactosidase, a secreted GH42 β -galactosidase, an arylsulfatase, an exported protein of unknown function, and a putative λ -carrageenase. λ -carrageenase cleave λ -carrageenan at the β -(1,4) linkage (121). It is likely that these proteins are playing a role in the degradation of these sulfated galactans, with the secreted β -galactosidase breaking bonds in carrageenan oligomer backbones. The degradation products generated from the λ -carrageenase and can potentially be imported with the ABC transporter, and the α -1,3-galactosidase would break the bond in the neocarrobiose, freeing the β -D-galactose-2-sulfate and D-galactose-2-6-sulfate, as λ -carrageen does not have 3,6-anhydro-bridges (121). The arylsulfatase could potentially remove the sulfurs from both monomers, and the D-galactoses can be fermented. We identified two more putative λ -carrageenases clustered together in the genome.

We identified a GH16 that is a putative κ -carrageenase, which also cleaves at the β -(1,4) linkage of κ -carrageenan(122), and this GH16 has a carbohydrate binding module (CBM) family 9 domain as well. This gene is adjacent to three more secreted CBM9 domain proteins, as well as a GH43 that is not secreted. These CBM9 proteins also have transmembrane domains, so they may function to attach the cell to the hemicellulose fibers of the red algae they are degrading. κ -carrageenan contains 3,6 anhydro-D-galactose monomers, which are released from degradation of the κ -neocarrabiose. We confirmed the presence of the already identified pathway for the assimilation of the 3,6 anhydro-D-galactose monomer found in (123), which includes the terminal steps of the DeLay-Doudoroff pathway. What was not mentioned in (123) is that the operon also contains 3 putative GH2 β -galactosidases, 4 sulfatases, a putative GH110 α -1,3-galactosidase, a putative GH127 β -L-arabinofuranosidase, and a CUT1 family carbohydrate binding protein. The β -galactosidases may also be involved in the processing of the imported carrageenan oligomers (both κ and λ), by degrading the β -(1,4) and α -(1,3) linkages, and removing the sulfates. Further we identified three operons in which β -galactosidases are present with sulfatases. These are likely involved in the intracellular degradation of the sulfated compounds present in red algae. The genome encodes for 30 sulfatases, with many sulfatase genes found near genes for CAZymes, an organizational theme found in other algae degrading bacteria (115, 124). Our findings suggest that type B is equipped with the abilities to degrade a diverse array of sulfated galactans found in the cell walls of red algae.

‘*Ca. Epulopiscium viviparous*’ may metabolize other complex polysaccharides found in the cell walls of red algae

The cell walls of red algae also contain microfibrils of cellulose, glucomannans, hemicellulose, and xylans (18, 112, 116, 125, 126). We were unable to detect any cellulases in the

type B genome for cellulose degradation. For the potential degradation of mannan, we were able to detect a putative secreted GH26 β -mannanase, which likely acts as an exo-mannanase or endo-mannanase, releasing mannobiose (116, 127). The genome also encodes for a cytoplasmic GH26 β -mannanase, as well as 2 GH2 β -mannosidases, and a β -(1,4)-mannooligosaccharide phosphorylase. It is likely that the extracellular mannobioses are imported in the cell where they are processed by the intracellular β -mannanase and β -mannosidases, to generate D-mannose, as well as by the β -(1,4)-mannooligosaccharide phosphorylase to generate D-mannose-1-phosphate. The β -(1,4)-mannooligosaccharide phosphorylase is located in an operon with and CUT1 ABC transport system likely specific for the mannan oligosaccharides, as well as a GH38 α -mannosidase. We were unable to detect any secreted glucanases to degrade the glucomannan components, but there are 3 putative cytoplasmic GH3 β -glucosidases to remove D-glucose from potentially imported oligomers.

The xylans found in red algae can have methyl esters and can have either β -(1,3) or β -(1,4) xylose linkages or a combination in their backbone, with glucosides and galactosides attached to the repeating xyloses (126). Arabinoses, acetyl esters, and sugar alcohols can also be found on xylans. Xylan degrading bacteria utilize a diverse array of enzymes including endo-xylanases, xylosidases, galactosidases, glucuronidases, arabinofuranosidases, glucosidases, and esterases to degrade this polysaccharide (116). We did not detect any secreted xylanases or xylosidases to degrade the xylan backbone, but did detect one putative cytoplasmic GH43 β -xylosidase and two other cytoplasmic GH43 enzymes that may function as a 1,5- α -arabinosidase, arabinofuranosidase, or a β -(1,3) xylanase. It is likely that other community members degrade long xylan chains into oligomers that type B can import and degrade releasing, xylose and arabinose. The genome also encodes for a cytoplasmic GH31 α -xylosidase and α -glucosidase, as well as 10 cytoplasmic GH2

β -galactosidases, which are part of the xylan degradation repertoire in *S. degradans* 2-40, and may be involved in oligomer degradation in type B (116). The genome also encodes for an exported GH53 protein which may function as an arabinogalactan endo β -(1,4)-galactosidase for removing galactose from the side chains (116). Included in the genome are 6 CE1 esterases, with four of which are cytoplasmic and two being exported, one cytoplasmic CE2 esterase, and four CE4 esterases, with one of which has a secretion signal. It is possible that some of these esterases degrade xylan (116). The genome also encodes a secreted GH16 with a CBM9 domain found in xylanases. It is possible that this hydrolase is able to degrade glucans or galactans that are found on red algae xylans (126). We also identified 2 genes adjacent to one another that encode secreted proteins with CBM6 domains resembling those found in xylanases, however, we did not observe a hydrolase domain in these proteins. One has a transmembrane domain, so it may function to attach xylan or xylan oligosaccharides to the cell surface. It is plausible that these enzymes lack a characterized hydrolase or lyase domain, and do function as novel xylanases. We did detect a CUT1 ABC transporter substrate binding protein with a CBM35 domain, making it likely that it is involved in the uptake of xylooligosaccharides. We detected 3 other CUT1 substrate binding proteins that are all putatively involved in the import of xylooligosaccharides present in the gut.

The type B genome encodes other CAZymes originally found in the genome of *Z. glactanivorans* including, 16 GH16 proteins, GH29 fucosidases, GH28 polygalacturonases, and GH74 proteins (115). All of these may play a role in red algal polysaccharide degradation. The genome content suggests that type B is equipped with the ability to degrade components of these structural polysaccharides.

‘*Ca. Epulopiscium viviparous*’ has the genomic potential to degrade xyloglucan and pectin, constituents of green algal cell walls

Naso tonganus feeds on green algae as well, including *Caulerpa* being observed in its diet (111). Like some red algae, the cell walls of *Caulerpa* contain β -(1,3)-xylan (128), while other marine green algae have mannans, pectins, arabinogalactan, xyloglucan, and/or cellulose in their cell walls (129-132). The genomic potential for type B to degrade xylans and mannans were already discussed, however, it appears as if type B is able to degrade the polygalacturon, pectin. We identified an operon containing a putative exported PL1 pectin lyase, 1 cytoplasmic PL9 pectate lyases, 2 exported and 1 cytoplasmic CE8 pectin methylesterases, and 2 exported PL9 pectate lyases. The PL9 lyases may serve as exopectate lyases at the non-reducing end of the polymer, and PL1 endo acting (116). We identified 2 more putative exported PL9 exopectate lyases, with one in an operon containing a CUT1 ABC transport binding protein and an FAD dependent oxidoreductase with a CBM32 domain, which has been found in proteins that bind galacturonic acid (133). Since we were unable to locate a 5-keto 4-deoxyuronate isomerase homolog to metabolize the 4-deoxy-L-threo-5-hexosulose-uronate generated from the PL9s, this oxidoreductase may serve in novel manner in the metabolism of this compound. Further we identified two cytoplasmic GH28 exo-polygalacturonases adjacent to each other in the genome. These enzymes may function to produce digalacturonic acid from larger oligomers (133), or generate D-galacturonic acid from oligomers. We identified two putative CUT1 ABC transporter substrate binding proteins that may be involved in the uptake of galacturon or polygalacturon in the genome.

Xyloglucan, a cell wall component found in some green seaweeds (132) is a complex polysaccharide made up of a β -D-(1,4)-glucan backbone in repeating units of 4, with xylose

attached to multiple glucoses in a α -D-(1,6) linkage. In the hemicellulose of land plants, xyloglucans can also include galactose, arabinose, or fucose attached through α -linkages in their side chains, and bacteria can degrade xyloglucan through the production of GH74 endo-xyloglucanases (134-137). These GH74 endoglucanases cleave the β -D-(1,4) glucan backbone generally at an unbranched glucose, producing different oligosaccharides, and can also cleave β -D-cellobioside and triside oligomers (135). The type B genome encodes for 15 putative GH74 proteins, 3 which are cytoplasmic and the rest appear to be exported. It is likely that some of these GH74 enzymes degrade the xyloglucan and its oligomers present in green algae. The genome also encodes for a cytoplasmic GH31 α -xylosidase to remove the xylose from the glucose backbone, as well as a GH35 β -galactosidase to remove any potential galactoses in the side chains. The genome encodes for a GH133 α -(1,6)-glucosidase to remove the glucoses from the side chains as well. The intracellular GH74 hydrolases may be generating glucose monomers, but this may also be done by the β -glucosidases present in the genome. We detected a putative *xyIP* transport system for the uptake of isoprimeverose, a breakdown product of xyloglucan. It is plausible that some of the GH74s are serving other functions, as 2 proteins have a CBM16 domain, which binds to mannans and cellulose. Further one of the GH74 domains is in a large 3,240aa protein, which also has a GH16 domain and CBM16 and CBM9 domain. This might serve instead as an endoglucanase breaking down glucans with its 2 GH domains. Two of the GH74 proteins could potentially act on β -D-(1,4)-glucomannans, as they are found in gene cluster containing a β -mannosidase.

Marine green algae produce multiple types of polysaccharide storage molecules and these can be sulfated. These include arabinans, galactans, arabinogalactans, amylose-type polysaccharides, and rhamnogalactans (130, 131). The 2 secreted G42 hydrolases may also be functioning to break down the β -(1,4) linkages in some of these galactans. It is possible that one

functions as a β -galactosidase and the other as an α -L-arabinase. The secreted GH53 endo β -(1,4)-galactanase, could also be acting on these storage galactans and arabinogalactans. In *Polaribacter* sp., the expression of GH74 hydrolases is upregulated during growth on chondroitin sulfate, making it possible that some of the GH74 hydrolases play a role in breaking down sulfated compounds of similar structures (124). A *Caulerpa* species produces a soluble glucan composed of a β -(1,3)-linkage glucose backbone, reminiscent of laminarin (see next section), with α -(1,4)-glucan linkages as side chains (138). The genome encodes for a secreted laminarinase that may function to degrade the β -(1,3) linkages in these compounds. It is plausible that in the large protein containing both a GH74 and GH16 domain, the GH16 domain is active against this compound. Some green algae produce inulin-type polyfructans as a storage molecule (139, 140). The type B genome encodes for a cytoplasmic GH32 β -fructosidase to degrade potential fructan oligosaccharides present in the diet, but we were unable to locate any specific transporter for the uptake of these compounds. Our findings suggest that type B is equipped with the ability to degrade certain complex polysaccharides like xyloglucan and pectin found in the cell walls of green algae, as well as their cytoplasmic storage molecules.

‘*Ca. Epulopiscium viviparous*’ encodes for the ability to degrade polysaccharides of brown algae

Brown algae is not known to be a major component of the diet of *N. tonganus*, however, brown thallate and filamentous algae can be found in their gut contents (111). The cell walls of brown algae contain alginate, a polysaccharide composed of repeating β -D-mannuronate and α -L-guluronate, and cellulose as structural fibrils (141). Many brown algal species also produce a storage molecule laminarin, which is a β -(1,3)-glucan with some β -(1,6)-glucose-linked side chains as well (142). We detected an operon containing a putative oligoalginate specific ABC

transport system, a cytoplasmic PL9 lyase, a secreted PL9 lyase, and a secreted GH16 laminarinase. Although no alginate lyases have been identified in the PL9 family, we hypothesize that based on the genomic context, these 2 lyases are alginate lyases, with the extracellular one acting on alginate, and the cytoplasmic lyase degrading oligoalginate, imported by the ABC transporter. We also detected a putative 4-deoxy-L-erythro-5-hexoseulose uronic acid (DEH) reductase to convert the DEH generated by the oligoalginate lyase to 2-keto-3-deoxy-D-gluconate (142). Further, the secreted laminarinase can generate laminari-oligosaccharides which may be imported through an unidentified transporter. The 2 GH3 β -glucosidases could potentially generate glucose from the oligosaccharides. The genome also encodes for a putative cytoplasmic GH30 β -(1,6)-glucanase to digest the side chain bonds. We detected another CUT1 ABC transport system likely involved in the uptake of alginate as well. These findings suggest that type B is competent to breakdown polysaccharides from brown algae.

Table 5. Polysaccharide utilization abilities of ‘*Ca. Epulopiscium viviparous*’

Algal polysaccharides	Genes for degradation	Genes for degradation of oligosaccharides
Ulvan	-	-
Xyloglucan	+	+
β -mannan	+	+
Glucomannan	+ [†]	+ [†]
Xylan	-	+
Cellulose	-	+
Laminarin	+	+
Alginic acid	+ [†]	+ [†]
Agar	+	+
Porphyran	+	+
κ -carrageenan	+	+
ι -carrageenan	-	-
λ -carrageenan	+	+
Pectin	+	+
Starch	+	+

[†] likely has the ability

‘*Ca. Epulopiscium viviparous*’ can assimilate a range of mono and disaccharides found in the host’s diet

Due to the complex carbohydrates type B degrades, it likely encounters multiple monosaccharides including D-glucose, D-mannose, D and L-galactose, D-xylose, L-arabinose, and L-fucose. Their metabolism of these compounds is summarized in Figure 3. We detected genes for a phosphotransferase (PTS) uptake system specific for glucose, as well as *mglABC* for the uptake of galactose, and *fucP*, which encodes a fucose permease. Further, we identified a *chvE* ABC transporter homolog for the uptake of multiple sugars including xylose, galactose, glucose, arabinose, and fucose. The genome also encodes for a PTS system for taking up fructose, as well as two CUT2 family ribose specific ABC transport systems. We were unable to detect any specific system for the uptake of mannose. It is possible that the FruA PTS may take up mannose at a lower affinity (143), or that the ABC transport system identified next to mannooligosaccharide utilization genes may be able to take up mannose monomers as well. The genome encodes for three putative Glycoside-Pentoside-Hexuronide/Cation (GPH) Symporter Family major facilitator superfamily sugar transporters, with one for the uptake of unspecified sugars. Mannose could potentially be a substrate for this transporter.

Glucose-6-phosphate generated from the glucose-specific PTS can be assimilated directly in the EMP pathway, while glucokinase can generate glucose-6-phosphate to assimilate non-phosphorylated glucose. D-Galactose can be assimilated through the Leloir pathway, generating D-glucose-6-phosphate. The fructose-1-phosphate generate from importation by FruA can be assimilated into the EMP pathway via FruK. The genome also encodes for a fructokinase to generate fructose-6-phosphate from D-fructose present in the cytoplasm. We were unable to detect a putative mannokinase to generate mannose-6-phosphate from the mannose generated from the

degradation of mannooligosaccharides. The glucokinase present in the genome is a member of the Repressor, Open reading frame, Kinase (ROK) family of kinases, which includes confirmed glucomannokinases (144). It is possible that this kinase, and/or the other ROK kinase present in the genome can function as a mannose kinase. The genome also encodes a phosphomanno/glucomutase to convert the mannose-1-phosphate generated from β -(1,4)-mannooligosaccharide phosphorylase to mannose-6-phosphate. *manA* encodes the ability to convert mannose-6-phosphate to fructose-6-phosphate, where it can funnel into the EMP pathway. The type B genome encodes for a complete pathway to degrade L-fucose to Dihydroxyacetone phosphate (DHAP), and (*S*)-lactaldehyde, which can be converted by *fucO* lactaldehyde reductase, to generate propane-1,2-diol, which can be excreted from the cell. D-xylose can be assimilated into the PPP through the isomerization to D-xylulose with XylA, and subsequent phosphorylation to D-xylulose-5-phosphate by XylB. L-arabinose can also be assimilated through the PPP as the genome putatively encodes for *araABD*, to convert L-arabinose into D-xylulose-5-phosphate. The genome encodes for ribokinase to phosphorylate the imported D-riboses to α -D-ribose-5-phosphate for assimilation into the PPP.

The genome encodes for 38 different CUT1 family carbohydrate ABC transporter substrate-binding proteins, which generally are involved in the uptake of di- and oligosaccharides (145). Of the 38, we were able to ascribe putative substrates to 15 of them, most of which were mentioned in the previous sections. We uncovered a putative maltose/trehalose CUT1 ABC transport system, as well as 2 maltose PTS systems, and a maltose specific CUT1 ABC transport system. It is likely that trehalose can also enter the cell through the glucose family PTS system. Red and green algae produce starch as storage polysaccharides (146), and these are likely degraded by other *N. tonganus* gut community members generating maltose. Trehalose is a disaccharide

generated in red, green, and brown algae that serves as an osmolyte and a storage molecule (147-149). The maltose-6-phosphate and trehalose-6-phosphate that enter through the PTS systems can be cleaved to a D-glucose and D-glucose-6-phosphate by the 2 6-phospho- α -glucosidases present in the genome. The non-phosphorylated maltose imported through the ABC transporters can be cleaved to 2 D-glucose molecules by the GH31 α -glucosidase present in the genome, while non-phosphorylated trehalose can be cleaved to 2 D-glucose molecules by a GH37 trehalase present in the genome. The genome also includes *sacP* a sucrose specific PTS system. Sucrose is a storage molecule and osmolyte in green marine macroalgae (149), and type B can potentially degrade the sucrose-6-phosphate it imports with a sucrose-6-phosphate hydrolase. We also detected a CUT1 ABC transport system putatively specific for stachyose and raffinose, which is produced by some green algae after cold shock (150). Enzymes to degrade both raffinose and stachyose appear to be encoded for in the genome due to the presence of the GH36 α -galactosidase, and GH31 α -glucosidase in the genome.

‘*Ca. Epulopiscium viviparous*’ has the potential to assimilate hexuronic acids found in the host’s algal diet.

D-Galacturonic and D-glucuronic acid are hexuronic acids found in green (129) and brown algae respectively (140, 151). Type B does not encode for a complete Entner-Doudoroff (ED) pathway, but does contain the terminal steps where KDPG is converted to pyruvate and D-glyceraldehyde-3-phosphate (G3P) by KDPG aldolase. The genome contains *uxaC*, *uxaB*, *uxaA* and *kdgK*, all of which are necessary for the 4 steps to convert D-galacturonate to KDPG. Aside from converting D-galacturonate to D-tagaturonate, UxaC, also functions on D-glucuronate, converting it to D-fructuronate. Genes for UxaB and UxaA are also present, and along with KdgK, to convert D-fructuronate to KDPG. We identified 4 GntP family transporters coded for in the

genome, with one annotated as a glucuronide transporter. The genome also encodes for a cytoplasmic GH2 β -glucuronidase to free D-glucuronate from imported brown algal D-glucuronides, allowing for their metabolism. One of the other GntP transporters looks to be specific for gluconate. There are multiple sugar kinases present in the genome and one may be able to phosphorylate D-gluconate. The genome encodes for 6-phospho-gluconate dehydratase to generate KDPG. GntP transporters serve as uptake systems for D-fructuronate in *E. coli* (152), making it possible that the other two transporters are involved in the uptake of hexuronates though they are not located in hexuronate utilization operons. The GPH transporter may also be able to import the hexuronates present in the algal diet of the surgeonfish host.

The ‘*Ca. Epulopiscium viviparous*’ genome encodes for the ability to degrade low molecular weight compounds found in the host’s algal diet.

Different species of marine algae produce polyols as storage molecules, osmololytes, and antioxidants. In some species of red algae, D-galactitol and D-sorbitol are produced (153), and these two compounds can be imported by the same galactitol PTS system, which is present in the genome, however, this PTS imports D-sorbitol with low affinity. A gene for galactitol-1-phosphate dehydrogenase is present in the genome to oxidize D-galactitol-1-phosphate to tagatose-6-phosphate, which can be converted to DHAP and D-glyceraldehyde-3-phosphate in two steps. The genome encodes for a putative D-sorbitol dehydrogenase, however, it is unclear if it can use D-sorbitol-6-phosphate as a substrate, but D-sorbitol can be converted to D-fructose through this enzyme. In brown algae, D-mannitol is the main product of photosynthesis (147), but can also be found in a few species of red and green algae (154). The PTS system specific for D-mannitol uptake MltA, is coded for in the genome, and the imported mannitol-1-phosphate can be oxidized to fructose-6-phosphate by mannitol-1-phosphate-dehydrogenase. The type B genome also

encodes for a complete pathway to degrade the polyol myo-inositol to DHAP and acetyl-CoA, however, it is unclear how this compound enters the cell.

Red algal species accumulate low molecular weight photosynthetic heteroside products composed of a monosaccharide and glycerol or glycerate and these compounds include digeneaside, floridoside, and isofloridoside (155). Type B encodes for the PTS transport system MngA, for the uptake of digeneaside (2-O- α -mannosyl-D-glycerate). *mngA* is in an operon with *mngB* and *glxK*, where MngB, an α -mannosidase, can cleave digeneaside, freeing the mannose-6-phosphate. GlxK is a glycerate kinase that can phosphorylates D-glycerate to 2-phospho-D-glycerate, which can enter into the EMP pathway. It is unclear if type B can utilize floridoside or isofloridoside as a carbon source. Red algae as well as green and brown algae also accumulate L-ascorbate, as it serves as an antioxidant (139). The type B genome includes 2 L-ascorbate specific PTS system, but does not include a complete pathway for L-ascorbate degradation. Located in an operon with one of the PTS systems are *ulaFEG*, which perform the first, third, and fourth steps in converting L-ascorbate-6-phosphate to D-xylulose-5-phosphate. However, the decarboxylase UlaD is not encoded in the genome. There are two genes for conserved proteins of unknown function located between *ulaE* and the PTS system genes, so it is possible that one of them performs the decarboxylation step, but there are no known protein family domains in these proteins to imply a function.

Table 6. Simple carbohydrate utilization abilities of ‘*Ca. Epulopiscium viviparous*’

Carbohydrate	Genes for assimilation	Carbohydrate	Genes for assimilation
3,6-anhydro-D-galactose	+	D-mannose	+
3,6-anhydro-L-galactose	+	D-sorbitol	+
D-fructose	+	D-xylose	+
D-fructuronate	+	L-arabinose	+
D-galactitol	+	L-fucose	+
D-galactose	+	L-galactose	+
D-galacturonate	+	L-rhamnose	-
D-gluconate	+	maltose	+
D-glucose	+	Mannitol	+
D-glucuronate	+	Sucrose	+
digenaside	+	trehalose	+

‘*Ca. Epulopiscium viviparous*’ can produce all 20 amino acids and assimilate ammonia

The ‘*Ca. Epulopiscium viviparous*’ genome likely encodes for the ability to produce all twenty proteinogenic amino acids, as nearly all the biosynthetic pathways are complete (summarized in Figure 3). The genome encodes for aspartate aminotransferase to make L-aspartate from oxaloacetate. It can be converted to L-asparagine and L-homoserine, which serves as a precursor to make both L-threonine and L-methionine. It has the capacity to synthesize L-methionine in a more direct route bypassing the need to make homocysteine, as the genome encodes for a homoserine O-acetyltransferase to generate O-acetylhomoserine, which can then be directly converted to L-methionine with O-acetylhomoserine thiolase, using H₂S or methanethiol for sulfurylation (156). We discovered a putative *mtsTUV* operon that is utilized by Firmicutes to import methionine precursors for biosynthesis (157). Threonine aldolase is present in the genome and can be used to convert L-threonine into glycine. Further a putative glycine-Na⁺ symporter was identified for the uptake of glycine, allowing them to obtain it from exogenous sources. From L-threonine, there is a complete pathway to produce L-isoleucine. L-aspartate is also an intermediate for the production of L-lysine, and the type B genome is able to make L-lysine from L-aspartate

through the DAP pathway variant in which DAP aminotransferase is used to make L,L-DAP directly (158). This method of making L-lysine is rather rare and is typically found in cyanobacteria (159), and a number of human pathogens (160), but some Firmicutes utilize this pathway as well (161). In Firmicutes, this enzyme is typically co-occurring with DAP dehydrogenase, giving them two options for L-lysine biosynthesis, but surprisingly, this gene is not present (160, 161).

The '*Ca. Eupulopiscium viviparous*' genome encodes for the enzymes required to convert pyruvate to 2-oxoisovalerate to make both remaining branched chain amino acids (BCAAs) L-valine and L-leucine. The *livJHMGF* operon is also present in the genome, and this is an ABC transport system for BCAA import (162). The genome encodes for a complete pathway to make L-histidine from D-ribose-5-phosphate, and chorismate from another PPP intermediate, D-erythrose-4-phosphate. To make chorismate, both shikimate dehydrogenase and shikimate kinase are necessary enzymes, and in the genome, the two genes are fused together. Chorismate can then be converted to prephenate via chorismate mutase, and from this compound L-tyrosine and L-phenylalanine can be made. The complete pathway is also present to convert chorismate to L-tryptophan.

From the glycolysis intermediate G3P, there is the capacity to make 3-phospho-L-serine, however, a gene for phosphoserine phosphatase is not present in the genome. We were able to discover three HAD family hydrolases, of which phosphoserine phosphatase is a member (163), and it is possible that one of the identified HAD hydrolases performs this function. L-serine is a critical amino acid as it serves as an intermediate to make others. The genome encodes for the two step process to make L-cysteine from serine, and also contains the enzyme cysteine desulfurase, which removes the sulfur group from cysteine to generate L-alanine. To offset its need to produce

L-cysteine, the genome also encodes the TcyP high affinity L-cystine/H⁺ symporter (164). Further, a putative SteT transporter gene was identified, which in *Bacillus subtilis*, is involved in the exchange of serine and threonine, and with a lower affinity, the exchange with aromatic amino acids (165, 166). As type B has a complete pathway for making L-threonine and the aromatic amino acids, it is possible to exchange them for the uptake of L-serine.

The type B genome provides it with the capacity to assimilate ammonia. Encoded is an NH₃ transporter for the uptake of NH₃ as well as an NADPH-specific glutamate dehydrogenase. This enzyme produces L-glutamate from 2-oxoglutarate and NH₃. Further, NH₃ can be assimilated into L-glutamine using glutamine synthetase. L-glutamine and 2-oxoglutarate can be converted back to two molecules of L-glutamate through glutamine oxoglutarate aminotransferase. From L-glutamate, the complete pathway to make L-proline is also present. L-arginine can also be synthesized from L-glutamate through the acetyl cycle. The genome encodes for the bifunctional ornithine acetyltransferase, which recycles an acetyl group from L-acetyl ornithine to generate L-acetyl glutamate, producing L-ornithine, which is further converted to L-arginine (167). Multiple PAAT ABC amino acid transporter systems were also identified. Based on sequence homology, one system is likely involved in the transport of basic amino acids including L-ornithine and octopine (168). The other ABC system is putatively ArtPQM, which is specific for the uptake of L-arginine (169). However, in the genome, the genes are present but not all are in the same operon, as *artPQ* are adjacent to one another. A putative Na⁺/dicarboxylate symporter was also discovered, and it may function in the uptake of both L-aspartate and L-glutamate. Due to its likely ability to produce all 20 amino acids it is plausible that type B is functioning to provide amino acids to its host.

PutP is a Na⁺/proline symporter and two copies of *putP* are found adjacent to one another in the genome. This permease is used by Enterobacteriaceae to import proline to use as a source of carbon and nitrogen via proline dehydrogenase (170, 171). Proline dehydrogenase is not present in the genome, making it unlikely that type B degrades L-proline in this fashion. Some Clostridia are able to ferment amino acids in Stickland reactions, in which one acts as a donor and another acts as an electron acceptor (172), and D-proline can serve as an acceptor. In *Peptoclostridium difficile*, this requires proline racemase and proline reductase (173), both of which are not in the type B genome, making it likely the permeases import L-proline for anabolic purposes. We were unable to detect a glycine reductase (174) or *ldhA* and *hadAIBC*, all involved in classic Stickland reduction reactions (172), making it unlikely that type B is conserving energy through Stickland reactions. Both subunits of serine dehydratase are encoded for, providing the potential to degrade L-serine to pyruvate. There is also the potential to generate ATP by degrading L-threonine to propionate, as the genome includes all five genes involved in this process. Unlike in some other bacteria, the genes are not clustered in the same operon, but the genes for phosphate acetyltransferase and propionate kinase, which perform the two terminal steps are in an operon together (175).

‘*Ca. Epulopiscium viviparous*’ genome provides the capability to use urea as a nitrogen source

We were able to identify genes involved in the import and degradation of urea. Adjacent to the cluster is the *urtABCDE* operon encoding the urea-specific ABC transporter, is the *ureABCEFGH* operon encoding for the three genes required to make the γ , β , and α subunits for a functional urease, and the four urease accessory proteins, necessary for maturation and activation of the metalloenzyme (176). The enzyme catalyzes the hydrolysis of urea to CO₂ and NH₃, which

can then be assimilated by the cell into L-glutamate or L-glutamine. In ruminants, nitrogenous waste in the form of urea is recycled by the rumen microbiota. Urea generated by the host is the primary supply of nitrogen for these bacteria, which convert it to NH_3 via ureases. The NH_3 is incorporated into biomass, which is in turn used by the host (177). To a lesser extent this occurs in the intestinal tracts of monogastric mammals such as mice (177), where the microbiota functionally breaks down urea as it can only be detected in the intestines of germ-free animals (178). In fish nitrogenous waste excretion typically occurs through the gills via ammonia diffusion. Urea formation is usually limited to the elasmobranchs, holocephalans, and coelacanths, and not teleosts like *N. tonganus*. In a few species urea can serve as an osmolyte for osmoregulation in high pH environments (179) or as a cloaking device to prevent predation (180). It is unclear if *N. tongonus* is producing urea along with ammonia, if so, type B may be serving an analogous function as the rumen microbiota, where it degrades urea and incorporates the ammonia released into biomass, which can be utilized by the host. This is not unheard of in teleost fish as the microbiota of *Porichthys notatus* appears to degrade urea. Antibiotic treated fish have a higher appearance of urea within the intestine, than untreated controls (181).

The ‘*Ca. Epulopiscium viviparous*’ genome encodes for a cytoplasmic nitrate reductase that may be involved in energy conservation

Located in the genome is an operon that encodes for an NADH dependent nitrate reductase. The operon matches the *nar* operon found in *Clostridium perfringens*, as it contains the *narCBA*, which encode for the NADH oxidase, the electron transfer, and the catalytic subunits of the reductase, respectively (182). The operon also contains genes (*moaAC*, *mobAB*, and *moeAC*) for the synthesis of molybdenum-molybdopterin guanine dinucleotide complex, which serves as a cofactor for the nitrate reductase (182). The exact role of the nitrate reductase in type B is unclear.

The genome also encodes for a potential ABC nitrate transport system from the nitrate/sulfonate/bicarbonate ABC-like transporter family found in cyanobacteria for NO_3^- import (183). Typically cytoplasmic nitrate reductases are involved in assimilatory nitrate reduction and annotated as *nas* genes, as the case with *B. subtilis* (183), however, some Clostridia potentially utilize this system to conserve energy. In *C. perfringens*, an inhabitant of animal intestinal tracts, NO_3^- serves as an electron sink; electrons are removed from pyruvate through NifJ, the pyruvate:ferredoxin oxidoreductase, to either a flavodoxin or a rubredoxin. Electrons are then transferred to the nitrate reductase, and on to NO_3^- as the terminal electron acceptor, generating NO_2^- (184). Using NO_3^- as an electron sink allows for additional substrate level phosphorylation for more ATP generation (185), and may also play a role in redox balancing (186). The type B genome encodes for NifJ, a flavodoxin, and rubredoxin, so it is possible that it performs a similar process, though the nitrate reductase would have to also accept electrons from rubredoxin as well. If their nitrate reductase is performing this dissimilatory process, it can allow type B to conserve more energy and better balance its redox potential, which may be factors in allowing for type B to reach such a large size. Also using NADH as an electron donor for nitrate reduction may allow for more energetically efficient NAD^+ recycling.

Some strains of *C. perfringens* are able to further reduce the NO_2^- to ammonia for assimilation. We were unable to detect any clear assimilatory or dissimilatory nitrite reductases in the genome of type B, however, there is a gene for a NirB family protein present. NirB is the large subunit of the *E. coli* nitrate reductase, and together with NirD form the functional reductase. A *nirD* gene does not appear to be adjacent in the genome, making it highly unlikely the NirB family protein is a nitrite reductase. The gene downstream of the NirB family protein gene, does encode for a small 118 amino acid protein annotated as an NAD dependent molybdopterin oxidoreductase.

Due to the improbability of NO_2^- reduction to NH_3 and its subsequent assimilation, type B must have a way of handling this toxic compound. Some nitrate reducing bacteria contain NarK, an antiporter that exchanges intracellular NO_2^- with NO_3^- . This gene is not present in the genome, however, we were able to identify a putative *nirC* gene. In *E. coli*, NirC functions as an H^+/NO_2^- antiporter which can perform both nitrite efflux and import (187). This transport system could provide type B with a means to detoxify itself of NO_2^- .

The ‘*Ca. Epulopiscium viviparous*’ genome provides for the ability to synthesize a complement of the B vitamins

The ‘*Ca. Epulopiscium viviparous*’ genes that could contribute to the metabolism of B vitamins is summarized in Table 7. The type B genome provides the cell with the capability to make the active cofactor of vitamin B₆, pyridoxal 5’-phosphate (PLP). Encoded in the genome are the genes for the two subunits of the PLP synthase complex, *pdxST*, which catalyze the one step synthesis of PLP from L-glutamine, D-ribose-5-phosphate and G3P (188). We were unable to locate any PLP transporters or genes involved in B₆ salvage in the genome. The genome also provides a complete pathway to make riboflavin (B₂), from GTP and D-ribulose-5-phosphate. *ribF* encodes for a bifunctional flavokinase/FAD synthase, which can convert riboflavin to the cofactors FMN and FAD.

The genome also encodes for the complete anaerobic pathway to make cobalamin (B₁₂). The biosynthesis of cobalamin is a very complicated process, involving multiple steps. Type B is able to synthesize precorrin II from L-glutamate. In some Clostridia, CobA and HemeD, two enzymes involved in the synthesis of precorrin II are fused (189). However, the type B genome encodes for a tri-domain polypeptide which arises from a fusion of those two genes along with the deaminase *hemC*, which is also involved in precorrin II biosynthesis. In the anaerobic pathway,

cobalt is inserted early into the heme center, by the chelatase CbiX, which is present in the genome. The rest of the genes involved in generating cobyrinate are present in an operon, with a *cbiE*, *cbiJ*, and *cbiT* gene fusion. The rest of the genes to create B₁₂ are also found in the operon, except *cobO*, which is present in another location. Also present at the beginning of the B₁₂ operon is an ABC transport system for the uptake of cobalamin. It is not exactly clear how type B generates dimethylbenzimidazole (DMB) for B₁₂ production. In aerobes, BluB is present to convert FMNH₂ to it in an oxygen dependent manner (190), while some anaerobes generate it from 5-aminoimidazole ribotide (AIR) (191). We were unable to locate the genes involved in the conversion of AIR, making it unlikely that they generate DMB in this fashion. There are multiple nitroreductases present in the genome, and it is possible that one of them plays an analogous role to BluB to generate DMB. The genome also encodes *cbiT* an ECF transporter for α -ribazole, which can be converted to DMB (192), meaning that if they cannot synthesize this molecule, they can acquire it from the external environment. Taken together '*Ca. Epulopiscium viviparous*' may be functioning to provide B₂, B₆, and B₁₂ to *N. tonganus*.

Present in the genome are multiple energy coupling factor transporters for the uptake of the vitamins they are unable to synthesize from either the host diet, or other microbiota members. In bacteria, multiple vitamins or their precursors are imported using these ABC-transporter family systems (157) and they are composed of four proteins, one transmembrane protein to form a permease EcfT, two ATP-binding proteins for the energetics of transport (EcfAA'), and an S-component for binding the substrate (193). There are two known classes, type I, where all the components are in a single operon, or type II, where *ecfTAA'* are not in the same operon as the S-component, and multiple substrate binding proteins can use the same EcfTAA' system (193). The type B genome appears to encode for class II ECF transporters. We identified two sets, one in

which the three components were maintained in a single operon, and another with the two ATP binding protein genes in one operon, and the permease in a standalone operon. *ribU* is present in the genome for the uptake riboflavin (194) and CblT is an S-component protein like RibU (157). The genome is lacking a complete pathway to make pantothenate (B₅), as it lacks a gene to make β -alanine, and multiple enzymes to convert pyruvate to (R)-pantoate, but contains the ECF S-component *panT* for its uptake. From pantothenate, the genome encodes for the complete biosynthetic pathway to generate coenzyme A. The biosynthetic pathway to the carboxylase cofactor biotin (B₇) is incomplete, but the genome encodes for BioY the ECF biotin binding protein, giving the cell the ability to import it from the environment.

Thiamin (B₁), a cofactor for enzymes involved in the essential carbon transformations is made from the combination of two compounds, hydroxymethyl pyrimidine pyrophosphate (HMP-PP) and thiazole phosphate carboxylate (THZ-P) (195). Type B lacks the genomic capability to make both substrates, but does encode the genes necessary for thiamin salvage. *thiM* is present for the salvage of THZ, as it phosphorylates free THZ, and *thiD* and *thiE* are fused. They encode for HMP kinase and thiamin phosphate synthase which combines THZ-P and HMP-PP to form thiamin monophosphate (TMP) (195). The three genes are located in the same operon. Typically, bacteria utilize TMP kinase (ThiL) to phosphorylate TMP to generate the active cofactor thiamin pyrophosphate (TPP), however, we were unable to locate *thiL*. Plants, fungi, and a few species of bacteria utilize a different pathway to generate TPP, where they dephosphorylate TMP with a HAD family hydrolase to thiamin and generate TPP via thiamin pyrophosphokinase (TPK) (196). We were able to detect a putative TPK to generate TPP from thiamin, as well as HAD superfamily phosphatase of subfamily IA, which was demonstrated to be dephosphorylate TMP in other bacteria (196), suggesting that type B generates thiamin this way from salvage products. Typically

these HAD hydrolases are found in the vicinity of thiamin biosynthesis genes, but the one present in the genome was located at the beginning of a contig, not near any thiamin genes.

It is unclear how type B salvages THZ, as we were unable to locate the THZ substrate binding protein ThiW (157, 197). We were unable to locate the *ykoCDEF* operon, which is involved in the uptake of HMP in *B. subtilis* (198), but were able to detect a putative *cytX* permease for the import of the HMP moiety (197). A putative ECF transport substrate binding protein *thiT* gene was discovered for the uptake of thiamin. Also located in the *thiMDE* operon is *tenA* which encodes for the thiaminase II. This enzyme catalyzes the intracellular degradation of base-degraded thiamin to restore a free HMP moiety for thiamin salvage (199). This provides type B with another potential route to acquire free HMP. In *Bacillus* species, ThiXYZ serves as the transport system to import base-degraded thiamin, however, we were unable to locate the genes for this system (199). It is possible that base-degraded thiamin can also enter the cell through *cytX*.

Folate (B₉) is an essential cofactor that serves as a donor and acceptor for one carbon compounds in purine and amino acid biosynthesis, and is synthesized in a multistep process to make both the pterin and para-aminobenzoic acid (pABA) moieties (200). pABA (vitamin B₁₀) is synthesized from chorismate, and type B can generate 4-amino-deoxychorismate, but there is no gene encoding the lyase that removes pyruvate from 4-amino-deoxychorismate to form pABA. In the genome, there are three genes annotated as *abgT* which in *E. coli* encodes for a transporter of the folate metabolite pABA-L-glutamate, but also requires the AbgAB proteins which hydrolyzes the bond between L-glutamate and pABA, freeing it for use in folate synthesis (201). We were only able to locate a putative *abgB* gene, so it is possible that this aminohydrolase does not require both subunits in type B or does not act on this substrate. Further, recent research has shown that AbgT family proteins in other bacteria function as an efflux pump for sulfonamide antimetabolites

(202), so they might be serving this function instead. pABA can also freely diffuse through the membrane into bacterial cells (202). The genome is also missing multiple genes involved in the production of the pterin moiety of folate from GTP, as well as dihydropteroate synthase to incorporate pABA. The lack of the ability to incorporate pABA makes it likely that the AbgT transporters function as antimetabolite efflux pumps. We were able to identify a putative *folT* gene encoding S-binding protein for folate. Dihydrofolate reductase is present in the genome and it can convert the imported folate into the active cofactor 5,6,7,8 dihydrofolate (THF) in two steps. Polyglutamate synthase is also present in the genome for the polyglutamylation of THF. The genes performing THF transformations are also present.

Nicotinate (B_3) is used to make the essential cofactor NAD^+ . NAD^+ can be synthesized by bacteria from L-aspartate or L-typtophan, however, the type B genome does not encode for the genes to perform these biosynthetic steps (203). NAD^+ can also be salvaged from nicotinate or nicotinamide. The genome does not encode for the *de novo* synthesis of these vitamins (204), but does have complete salvage pathways to generate $NAD(P)^+$ from both of those compounds. How type B acquires exogenous B_3 remains elusive as we were unable to detect the *niaX* gene which encodes for a nicotinamide and nicotinate ECF substrate binding protein. *pnuC* is also not present in the genome, which is the transporter for nicotinamide riboside, an NAD^+ biosynthesis intermediate (204). This transporter is present in the genome of *C. lentocellum* DSM 5427. It is possible the inability to find this gene is due to a divergent transporter, or an assembly error. It is also possible that type B has an uncharacterized transporter for nicotinamide mononucleotide.

Table 7. B-vitamin biosynthesis and import ability of ‘*Ca. Epulopiscium viviparous*’

Vitamin	Synthesize	Import
Cobalamin	+	+
Folate	-	+
Nicotinate	-	-
p-Amino benzoic acid	-	-
Pantothenate	-	+
Pyridoxal 5'-phosphate	+	-
Riboflavin	+	+
Thiamin	+	+

* ability to synthesize from salvaged products *N. tonganus* cannot utilize

‘*Ca. Epulopiscium viviparous*’ has complete gene suites to produce purine and pyrimidine bases, but does not contain nucleoside diphosphate kinase in the genome

Bacteria utilize the PPP intermediate D-ribose-5 phosphate as the starting point to make purines and pyrimidines in a multistep process. In the first step, D-ribose-5 phosphate is converted to 5-phospho- α -D-ribose 1-diphosphate (PRPP). ‘*Ca. Epulopiscium viviparous*’ encodes all 11 steps to make the purine intermediate inosine monophosphate (IMP). *purD* and *purN*, which are necessary to make IMP are fused together in the *purLECFMDN* operon, but *purK* is not present in this operon. From IMP, type B has the potential to make AMP using PurA to generate an adenylosuccinate intermediate, which is then converted to adenosine monophosphate (AMP) via adenylosuccinate lyase. AMP can be phosphorylated to ADP via adenylate kinase. Nucleoside diphosphate kinase (*ndk*) is not present in the genome to convert ADP to ATP, but ADP can be converted to ATP through pyruvate kinase. To generate dATP for DNA polymerization, anaerobic nucleoside triphosphate reductase (NrdD) is present. *guaAB* are present and encode IMP dehydrogenase to generate xanthosine monophosphate, and GMP synthetase to generate guanosine monophosphate. Guanylate kinase is present to generate GDP and GTP can likely be generated by

pyruvate kinase. Multiple Mollicutes with complete genomes lack *ndk* as well, but their glycolytic kinases like pyruvate kinase and phosphoglycerate kinase can use other nucleotide and deoxynucleotide diphosphates as phosphate acceptors, with GDP, CDP, dTDP, and UDP all being converted to the triphosphate forms by pyruvate kinase (205). It is plausible that this is occurring in type B. dGTP can be generated by NrdD from GTP. There are two putative GTP pyrophosphokinases present in the genome to generate pppGpp, and an exopolyphosphatase to generate ppGpp.

To make the pyrimidines, PRPP is converted to orotidine-5 phosphate. To do this, bicarbonate must be converted to orotate in 5 steps, and the genes for this conversion are present. Orotate phosphoribosyltransferase combines orotate and PRPP to generate orotidine-5 phosphate. From this, uridine monophosphate (UMP) can be generated by PyrF, and uridylate kinase can generate UDP. UTP is likely generated from pyruvate kinase, or another set of glycolytic kinases that have less specificity for substrates. UTP can be interconverted to CTP through CTP synthase. NrdD can generate dCTP and dUTP, and the latter is converted to dUMP by dUTP pyrophosphatase, and then to dTMP by thymidylate synthase. There are two thymidylate kinases present in the genome to generate dTDP from dTMP, and pyruvate kinase likely generates dTTP.

Another possibility for generating nucleoside triphosphates is through the usage of polyphosphate kinase. In *Pseudomonas aeruginosa* and *E. coli* mutants lacking Ndk, nucleoside triphosphates could still be generated from the polyphosphate dependent polyphosphate kinase (Ppk) (206). In these species, polyphosphate kinase is localized to the membrane (206). Type B encodes *ppk*, but it is unclear if this localization occurs in type B. If this is the case, it may be more evidence that this kinase is involved in nucleotide synthesis as the DNA is localized at the membrane in type B (23, 29). This localization could make for more efficient DNA synthesis

during growth, reducing the transit time of dNTPs to reach DNA polymerase III. For the accumulation of polyphosphate, the genome encodes for a Na⁺/P_i dependent phosphate symporter. To our knowledge PEP carboxykinase has not been investigated in other organisms for its ability to utilize other NDP substrates, but it can be another prospective glycolytic kinase involved in NTP production. It is plausible that both the glycolytic kinase and polyphosphate kinase systems are being used to overcome the lack of an Ndk.

The ‘*Ca. Epulopiscium viviparous*’ genome is quite different from the ‘*Ca. Thiomargarita nelsonii*’ genome.

Although *Epulopiscium* spp. are the largest heterotrophic bacteria, they are not the largest known group of bacteria, this distinction belongs to members of the genus *Thiomargarita* which contains large colorless sulfur-oxidizing bacteria that can reach diameters of 800 µm (207). Like *Epulopiscium* spp. these large bacteria are also highly polyploid (67), with DNA located at the periphery of their cells. Due to the gigantic nature and large quantities of DNA found in both groups of organisms we sought to uncover the overarching genomic characteristics of large bacteria. Recently two draft genomes for ‘*Ca. Thiomargarita nelsonii*’ strains have been published, serving as a basis for comparison (Table 8). The ‘*Ca. Thiomargarita nelsonii*’ species group is composed of cells of different morphotypes, including chain-forming square shaped cells (Figure 6B) like ‘*Ca. Thiomargarita nelsonii* Thio36’, large spherical cells that aggregate in cube like assemblies, or free-living spherical cells attached to surfaces that produce spherical buds at the apical ends, like ‘*Ca. Thiomargarita nelsonii* Bud S10’ (208). Members of this species group can achieve sizes of up to 250 µm in diameter (208).

By comparing the genomic potential for energy conservation, it becomes apparent that there is not one metabolic strategy required to achieve these large sizes. As described in the

above sections, type B relies solely on fermentation to conserve energy. In contrast both '*Ca. Thiomargarita nelsonii*' strains apply a mixotrophic lifestyle in which they can grow lithotrophically on reduced sulfur compounds, but have the genomic potential for H₂ oxidation (209, 210), with Bud S10 having the potential for arsenite oxidation (210). The '*Ca. Thiomargarita nelsonii*' strains can respire aerobically, but can also use nitrate as a terminal electron acceptor which it can denitrify to N₂, with energy conserved through oxidative phosphorylation (209, 210). Type B does not have an electron transport system but does encode multiple redox proteins including flavodoxins, ferredoxins, and thioredoxins to potentially help balance redox potential. Like '*Ca. Thiomargarita nelsonii*', it also can use NO₃⁻ as an electron acceptor, though not in the typical respiratory sense. Like type B, both '*Ca. Thiomargarita nelsonii*' strains can potentially generate an SMF, but they achieve this through a different mechanism as they encode for the rnf complex (209, 210). Another major difference between these organisms is the autotrophic ability for '*Ca. Thiomargarita nelsonii*' strains to fix carbon via the reverse TCA cycle and the Calvin-Benson-Bassham cycle as both are encoded for in the genomes (209, 210). Type B lacks the genomic potential for this, and uses organic carbon for biomass. Similar to type B, '*Ca. Thiomargarita nelsonii*' also encodes for amino acid and peptide transporters, supporting it having a mixotrophic lifestyle (210).

A major difference between the two organisms is the presence of mobile genetic elements. The genome '*Ca. Thiomargarita nelsonii* Bud S10' is more plastic as it is littered with mobile genetic elements, which disrupt polymerases, transcriptional regulators, and transporters (210). The Bud S10 genome contains multiple group I introns, inteins, fdxN excision elements (site specific recombinases), and Miniature Inverted-repeat Transposable Elements, (MITEs), none of which were found in the type B genome. Bud S10 contains 55 other putative

transposase/insertion sequence (IS) elements, with 28 belonging to the TIGRfam family TIGR01784, with a PD-(D/E)XK_2 homing endonuclease domain (PFAM12784) (210). We were able to detect 2 genes in the type B genome that contained both of these domains, suggesting there are 2 transposases present in the genome. Bud S10 had 55 hits aligning to a consensus group II intron sequence, suggesting they were prevalent in the genome, and contains the PFAM00078 reverse transcriptase motif 62 times in its putative proteins (210). There were 2 detected group II introns found in the type B genome, one disrupting the gene for the fibronectin binding protein *yloA* (211), and adjacent to the group II intron was a group II-like reverse transcriptase/maturase domain protein. The other group II intron did not have the reverse transcriptase protein next to it, and it disrupted *dnaE3*. As in type B, *dnaE3* is found in *B. subtilis* along with *polC* (212), which encodes for the, replicative α subunit of DNA polymerase III and has 3' \rightarrow 5' exonuclease activity. In *B. subtilis* its DnaE3 functions to add a few dNTPs to the RNA primer before PolC takes over (213). Due to the group II intron disruption, the type B DnaE3 is likely not functional, suggesting that PolC performs this function. Despite the reduction of mobile elements found in type B in comparison to Bud S10, it is possible that polyploidy in both organisms functions to mitigate the effects of these mobile element disruptions, and there are low numbers of genomes with functional copies of these genes present. The group II intron associated reverse-transcriptase was the only one found in the type B genome. The high number of mobile elements in Bud S10 likely accounts for why 18% of the its genome is intergenic (210), whereas with the low level of mobile elements, the type B genome has a 92% coding density.

Motifs for metacaspases were a prominent feature in the Bud S10 genome. Metacaspases function in programmed cell death, but have other cellular functions as well (214). Due to the

type B life-cycle in which the mother cells undergo a programmed cell death for the emergence of daughter cells, it seemed plausible that metacaspases could be detected, however, no metacaspase were present in the genome. This aligns with what is found in *B. subtilis* as no metacaspases are present in its genome for the programmed cell death of the mother cell during sporulation.

Both type B and '*Ca. Thiomargarita nelsonii* Bud S10' have repetitive elements in their genomes. Type B has large blocks of imperfect direct repeats in multiples of 3 (typically 18 to 21bp long) within the genome (Figure 6). These blocks are present in each contig throughout the genome, but there tends to be concentrated areas of repeats within contigs (Figure 1). These repeats typically do not share the same sequences, and are found in intergenic regions, or stretch for distances that include small proteins of unknown function. A few of these repeated sequences can be found in some of the large genes as discussed above. Like other Beggiatoaceae, the heptamer TAACTGA is found as a direct or indirect repeat scattered throughout the Bud S10 genome. Although these repeated elements are not of the same structure as in type B, repetitive DNA elements may be necessary structural elements for these large polyploid bacteria. It is plausible that these repeat elements are recognized and bound by proteins that help maintain the DNA at the periphery of these organisms (207). Another shared characteristic is that all three genomes contain the genomic capabilities to accumulate polyphosphate. All 3 organisms encode for a polyphosphate kinase, an exopolyphosphatase, and Na^+/P_i co-transporters, with Bud S10 containing 2 more transport systems for P_i . Both type B and Bud S10 organisms encode for 2 putative *pfkA*, with one in type B being pyrophosphate dependent and at least one likely pyrophosphate dependent in Bud S10 (210). Further, Thio36 has a polyphosphate dependent glucokinase, and the type B genome encodes for a polyphosphate dependent NAD kinase (209). It

is possible that due their size and seemingly concomitant increase in energy requirement, these organisms may utilize polyphosphate as a phosphate donor in certain situations instead of ATP to help conserve intracellular ATP.

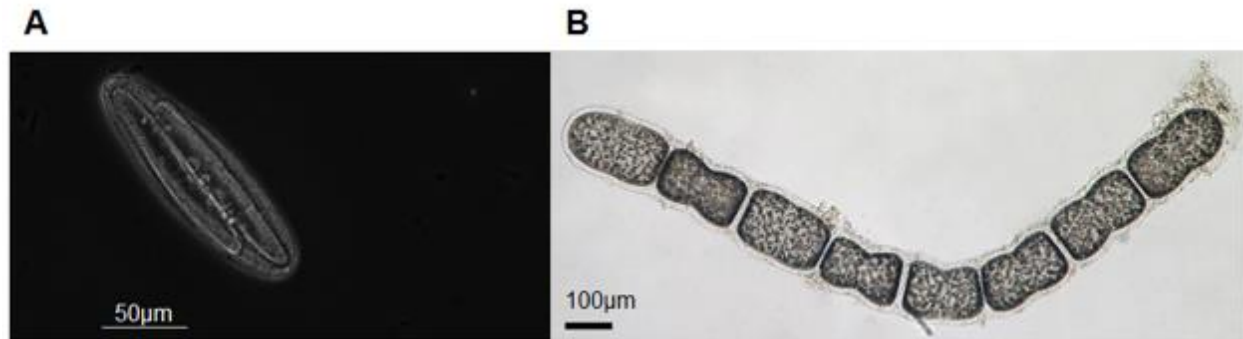


Figure 6. ‘*Ca. Epulopiscium viviparous*’ and ‘*Ca. Thiomargarita nelsonii*’ micrographs. (A) is a bright-field micrograph of a ‘*Ca. Epulopiscium viviparous*’ cell fixed in ethanol after removal from the gut contents of *N. tonganus*. Visible within the cells are intracellular daughter cells. (B) is a bright field micrograph of a filament of ‘*Ca. Thiomargarita nelsonii*’ cells, like ‘*Ca. Thiomargarita nelsonii* Thio36’. The image was taken and provided by Verena Salman-Carvalho.

AAATCTTCGAGAGCCGATGTGAAATCTTCGAGAGCCGATGTGAAATCTTCGAGAGCCGAT
GTGAAATCTTCGAGAGCCGATGTGAAATCTTCGAGAGCCGATGTGAAATCTTCGAGAGCC
GATGTGAAATCTTCGAGAGCCGATGTGAAATCTTCGAGAGCCGATGTGAAATCTTCGAGA
GCCGATGTGAAATCTTCGAGAGCCGATGTGAAATCTTCGAGAGCCGATGTGAAATCTTCG
AGAGCCGATGTGAAATCTTCGAGAGCCGATGTGAAATCTTCGAGAGCCGATGTGAAATCT
TCGAGAGCCGATGTGAAATCTTCGAGAGCCGATGTGAAATCTTCGAGAGCCGATGTGAAA
TCTTCGAGAGCCGATGTGAAATCTTCGAGAGCCGATGTGAAATCTTCGAGAGCCGATGTG
AAATCTTCGAGAGCCGATGTGAAATCTTCGAGAGCCGATGTGAAATCTTCGAGAGCCGAT
GTGAAATCTTCGAGAGCCGATGTGAAATCTTCGAGAGCCGATGTGAAATCTTCGAGAGCC
GATGTGAAATCTTCGAGAGCCGATGTGAAATCTTCGAGAGCCGATGTGAAATCTTCGAGA
GCCGATGTGAAATCTTCGAGAGCCGATGTGAAATCTTCGAGAGCCGATGTGAAATCTTCG
AGAGCCGATGTGAAATCTTCGAGAGCCGATGTGAAATCTTCGAGAGCCGATGTGAAATCT
TCGAGAGCCGATGTGAAATCTTCGAGAGCCGATGTGAAATCTTCGAGAGC

Figure 7. Example tandem repeat found in the ‘*Ca. Epulopiscium viviparous*’. Represented is the direct repeat found at the end of one of the contigs in the genome.

Table 8. Comparative table contrasting ‘*Ca. Epulopiscium viviparous*’ and both ‘*Ca. Thiomargarita nelsonii*’ genomes

Property	‘ <i>Ca. Epulopiscium viviparous</i> ’	‘ <i>Ca. Thiomargarita nelsonii</i> Thio36’	‘ <i>Ca. Thiomargarita nelsonii</i> Bud S10’
nucleotides	3.8 Mb	5.3 Mb	6.2 Mb
ORFs	2,714	7,596	7,525
coding percentage	92	72	82
tRNAs	54	23	46
GC%	38.08	42	41.3
genome completeness %	92	70	89.8
Typical cell size	200-300µm by 50-60µm		
Habitat	<i>Naso tonganus</i> gut symbiont	Sulfidic sediment surface water	Sulfidic sediment surface water
Energy conservation	fermentation	Lithotrophy	Lithotrophy
SMF generation	oxaloacetate decarboxylase	rnf complex	rnf complex
Nitrate reduction	dissimilatory - non-respiration	dissimilatory (denitrification) and assimilatory	dissimilatory (denitrification) and assimilatory
Carbon fixation	Do not fix carbon	CBB Cycle, Reverse TCA	CBB Cycle
Degree of polyploidy	10,000s of chromosome copies	Not known	Not known
Polyphosphate accumulation	+	?	+
reverse transcriptases	1	?	62
Presence of group I introns	-	?	+
Presence of group II introns	+ (2)	?	+ (57)
Presence of transposons/IS elements	+ (2)	?	+ (90)
Presence of MITEs	none detected	?	+
Presence of inteins	none detected	?	+
fdxN Excision Elements	none detected	?	+

5.5 DISCUSSION

Utilizing multiple sequencing technologies afforded us the ability to generate a high quality draft genome of '*Ca. Epulopiscium viviparous*.' This approach was beneficial for multiple reasons; first using long PacBio reads helped mitigate some of the assembly issues that would previously arise because of the repetitive nature of the genome. It also allowed for us to get a better estimate of the quantity of rRNA genes present in the genome. Secondly, using paired-end Illumina reads provided us with a more accurate subset of reads, allowing us to error correct our PacBio reads in this manner. It also provided us with paired-end information to do linkage mapping to bioinformatically join contigs, as well as remove potentially contaminating sequences since the genomic DNA was essentially a metagenomic sample enriched for type B. Third, having both sequencing technologies allowed us to utilize multiple assemblers, each with their own intrinsic advantages, providing us with an array of assemblies to work with. We found that Canu, part of the Celera assembler, provided the best assembly when LoRDEC-corrected PacBio reads served as the input, improving upon using PacBio self-corrected reads. The Illumina-only assemblies were not as complete as those containing PacBio reads, making PacBio necessary for a high quality genome. However, the assembly did have some flaws as there were multiple cases in which other assemblies did not have gaps that were present in this assembly. By comparing assemblies composed of different reads and generated by different assemblers, and finding shared sequences, we had the confidence to close these gaps. This approach allowed us to further correct multiple mis-assemblies present in the genome. Although not closed, the genome is estimated to be between 92-97% complete, and has provided us with great insight into the biology of type B. Thus this 3.28Mb genome broken into 7 contigs is a vast improvement upon the genome used by Miller and colleagues to investigate reproductive genes (27).

‘*Ca. Epulopiscium viviparous*’ has only been found in the GI tracts of *N. tonganus*, thus its maintenance in the gut is tantamount to its fitness and survival. Further, type B is preyed upon by the ciliate *B. jocularum* (72), therefore it needs to evade predation to survive in the gut. Type B encodes for giant surface proteins that may help protect itself from these predatory eukaryotes (71). Some of these proteins also include Ig fold domains found in adhesins. This may also be important for maintenance of positioning within the gut, especially when the mother cells become sessile as their daughter cells are readying to emerge. For the benefit of their daughter cells, the mother cells may use these large proteins to adhere to the mucous membrane or near the food bolus to prevent daughter cells from getting eliminated with feces and keeping them near the food supply. ‘*Ca. Epulopiscium viviparous*’ is also motile and encodes for 9 different flagellin genes. The diversity of flagellin genes allows for numerous variations of the flagella type B can produce, which may also be important for binding to food or the gut mucosal layer for maintaining position within the gut. If the flagella do play a role in type B binding, the variation in flagellin monomers within the flagella may be necessary for improving binding strength due to type B’s large size. Genomic evidence suggests type B is chemotactic which likely is necessary for maintaining position in the gut and avoiding predation by *B. jocularum*. The presence of genes involved in the production of cell surface glycopolymers may also play roles in evading ciliate predation and making the large cells sticky again for positioning within the gut, possibly also for attachment of sugars and polysaccharides for consumption.

With some surgeonfish species, diet appears to be a key driver in determining microbiota composition, and herbivorous surgeonfish have a high abundance of *Epulopiscium* spp. in their guts (3). There has not been a comprehensive examination of the *N. tonganus* microbiota community or the factors that shape it, so it is unclear to what extent ‘*Ca. Epulopiscium viviparous*’

dominates the gut community. *N. tonganus* is omnivorous and its diet is abundant in red and green algae, and to a much lesser extent filamentous brown algae (111). Based on genomic potential '*Ca. Epulopiscium viviparous*' appears to be geared for degrading and assimilating the diverse array of polysaccharides encountered in *N. tonganus*' diet. The genome is enriched for carbohydrate metabolism COGs, and over 5% of the genome is dedicated to CAZyme production, a value rare for bacteria, even for those that specialize in carbohydrate degradation (110). The diversity and abundance of glycoside hydrolases encoded for by the genome are similar to those observed in algal degrading free living bacteria (109, 110). Polysaccharides are the most diverse class of macromolecules in nature and contain various linkages and moieties. The diversity of CAZymes present in type B's genome likely allows for it to accommodate some of this diversity. The 12 different putative agarases from 3 distinct GH families is likely necessary to degrade the variations of agar found in different red algae. Further, it is possible that there are novel hydrolases and lyases present in the genome. Biochemical analyses on these putative GHs and PLs need to be conducted to better understand the substrates and functions of these enzymes. Type B may be reliant on other community members to break down certain polysaccharides, as no known cellulosome or secreted xylanases were uncovered in the genome, but intracellular enzymes to further degrade oligosaccharides were found. It is plausible that one of the reasons why type B cell sizes are so large is that they take up large pieces of polysaccharides and degrade them internally, while also degrading others externally. The mother cells also need to 'feed' the daughter cells as they develop, so large amounts of internal degradation may be a way of achieving this. This would have the added benefit of preventing other community members from having access to these oligomers, potentially giving type B a fitness advantage. The large size would be necessary to accommodate such a strategy.

This enrichment for polysaccharide degradation likely benefits the host. Type B can assimilate polysaccharides and sugar alcohols present in the diet that the host alone cannot. Due to its fermentative abilities, acetate is likely the main fermentation product produced by ‘*Ca. Epulopiscium viviparous*’. As type B appears to be the dominant community member in *N. tonganus*, this may account for why acetate is the most abundant short-chain fatty acid in *N. tonganus*’ gut. Peak acetate concentrations were found in the gut where type B resides and downstream of its residence (18). The acetate produced by type B can be absorbed by the gut (18) and contribute to the nutrition of *N. tonganus*. We were unable to detect genes involved in butyrate formation, and it only appears that type B can potentially generate propionate by converting threonine to 2-oxobutanoate then to propionyl-CoA and eventually to propionate. This metabolism is likely not occurring at high levels due to the importance of L-threonine and 2-oxobutanoate, and there is not much propionate found in the gut of *N. tonganus* (18). This conversion of recalcitrant material to usable energy for the host is reminiscent of what occurs in mammalian guts (214).

The nitrogen content of algal diets in marine organisms is not clear as there are complications with properly assessing it, with some studies suggesting that it may be low (19). ‘*Ca. Epulopiscium viviparous*’ may be able to supplement *N. tonganus*’ diet with amino acids and protein if dietary content is low. Type B has the genomic potential to make all 20 amino acids, and due to its complicated life cycle in which these large cells undergo programmed cell death and lyse in a diurnal fashion, they may be contributing an appreciable amount of protein to *N. tonganus*. The physiology of *N. tonganus* is not well understood and it is unclear if both NH_3 and urea are both nitrogenous waste products created by the fish, but type B has the genomic potential to incorporate both into protein, effectively recycling these waste products into a usable form. B-vitamins are another nutrient that fish intestinal microbiota may supply their host (15), and type B

has the ability to synthesize riboflavin, pyridoxal 5'-phosphate, and cobalamin *de novo*, as well as thiamin from scavenging the two moieties either from the algal diet or other microbiota members. Since animals are unable to synthesize their own B-vitamins, '*Ca. Epulopiscium viviparous*' may be a reservoir for those 4 for *N. tonganus*.

Fish microbiota also functions to stimulate the immune system and epithelial cell renewal (4). Type B is not a member of the juvenile *N. tonganus* microbiota, so they are not present as the gut and immune system develop. This may further account for why '*Ca. Epulopiscium viviparous*' encodes for 9 different flagellins, as this may be necessary to not trigger an immune response in the host during initial colonization. The production of cell surface glycopolymers may also serve this function, and there is also the possibility that type B produces a poly- γ -glutamate capsule as well. The ESX type VII secretion system encoded for in the genome may also function in interaction with the immune system of the host to prevent an immune response.

The '*Ca. Epulopiscium viviparous*' genome provides more insight into the energetics of this large heterotroph. The lack of genomic evidence for a respiratory electron transport chain suggests that type B is an obligate fermenter, but the genome suggests type B has two mechanisms to conserve more energy. It encodes for a dissimilatory cytoplasmic nitrate reductase that may accept electrons from pyruvate via NifJ, but can from NADH making NO_3^- an electron sink for redox balancing. The genome also encodes for the ability to convert PEP to oxaloacetate, as well as to ferment citrate to oxaloacetate. Oxaloacetate can then be decarboxylated to pyruvate via the oxaloacetate decarboxylase which couples this reaction with the extrusion of Na^+ , generating a SMF. The F_1F_0 ATPase appears to be Na^+ specific, and may have some unique structural components, due to the triplication of the AtpE domain, and the presence of this can enhance ATP production from just relying on substrate level phosphorylation. The genome also encodes for

multiple uncharacterized oxidoreductases, dehydrogenases, and other redox-reactive proteins that may be playing unique functions in energy conservation. These proteins need to be explored biochemically to understand the functions and substrates of these proteins.

The nutrient rich environment of the gut and diversity of carbohydrate degradation found within the genome likely allows for '*Ca. Epulopiscium viviparous*' to maintain its large cell size. Surprisingly the genome lacks *ndk*, and carbohydrate kinases like pyruvate and acetate kinase may function in making the NTPs and dNTPs necessary for RNA and DNA synthesis. Ultimately, all carbohydrate metabolism in type B funnels into the glycolytic EMP pathway and pyruvate kinase performs the terminal step in this pathway. If it and other glycolytic kinases like acetate kinase can utilize multiple nucleotide diphosphates as substrates, this would directly couple cell growth to central metabolism, as the cell would only be able to produce RNA and DNA when it is fermenting. This may be an energetic adaptation only allowing for replication during periods of high metabolic activity. This may help regulate the diurnal cycle observed in type B as their DNA and RNA synthesis would be beholden to host feeding for providing them with fermentable substrates. Daughter cells develop within the mother cell and replicate their genomes when the fish is active and feeding, and as they fill up the mother cell's cytoplasm, the mother cell's DNA begins to deteriorate (32). At these final stages of mother cell life, loss of metabolic activity and therefore the ability to make dNTPs may help account for the deterioration of DNA. Further it is possible that this a form of genome streamlining, where generating NTPs and dNTPS in this fashion is more energetically efficient than utilizing a designated kinase specific for only one purpose. If this is more energetically efficient, it may help explain how type B is able to be so polyploid and large. It can also possibly explain why terminally differentiated mother cells still synthesize DNA after

passing chromosomes to their progeny as they are still metabolically active and there may be a buildup dNTPs during fermentation.

To further offset the energetic constraints of being large and the costs of making a gene (67), a theme found in multiple biosynthetic pathways and carbohydrate transporters are fused genes. According to IMG, almost 6% of the protein coding genes are fused in the genome. This effectively increases the coding density and reduces the requirement for regulatory elements and intergenic spaces freeing up some energy (67). Gene fusions also likely improve the efficiency of a pathway, so along with reducing energy input this could reduce the number of proteins that have to localize to a certain location, as the domains are fused in one polypeptide. This adaptation may be a consequence of its large size, but could also be a form of genome streamlining found in symbionts, as its association with *N. tonganus* appears to be obligate. Fusions could also be a consequence of the high rates of recombination, in which genes fuse to prevent operons from being broken up. Type B also has the potential to accumulate polyphosphate and may use polyphosphate kinase to also generate NTPs and dNTPs. Further it has the ability to use polyphosphate as a donor instead of ATP in glycolysis as it has a polyphosphate dependent *pfkA*. This would help offset the ATP demand and conserve intracellular levels. The genomic landscape sets the foundation for understanding the energetics of these large organisms allows them to maintain such a large cell size and volume.

Lane and Martin argue that unlike protein synthesis in cells, which scales with cell volume, ATP synthesis scales with cell surface area due to respiration, and as a result larger bacterial cells are less energetically efficient (67). However, type B does not appear to be reliant on respiration for ATP synthesis, and consequently surface area may not be as large of a constraint for scaling up ATP synthesis. It may be that having a large volume is more necessary for massive amounts of

intracellular poly- and oligosaccharide fermentation to occur. Increased surface area would still benefit the cell in terms of ATP synthesis as it would provide more area for carbohydrate uptake and its subsequent degradation, resulting in more ATP production. ‘*Ca. Epulopiscium viviparous*’ membranes are heavily invaginated likely for this purpose, but if large polysaccharides and oligosaccharides can be imported and degraded internally, it may help mitigate the problem of the small surface area to volume ratio. It may be that through polyploidy and the reliance on fermentation for ATP synthesis through both substrate level phosphorylation and an ATP synthase, the energetics may be more favorable for a large cell size than through respiration. The large cell size may accommodate larger amounts of fermentation, as it would allow for the scaling up of protein production, including intracellular polysaccharide degradation, reciprocally providing enough energy to allow for large cell growth.

The ‘*Ca. Epulopiscium viviparous*’ draft genome is the first of the *Epulopiscium* spp. group and depicts the largest sequenced heterotroph. It has shed light on the organism’s biosynthetic and metabolic capabilities, its energetics, and how it may be interacting with its host. Further it provides a basis for performing more functional analyses on this unique organism as it will allow for more effective RNAseq studies and may help lead to the culturing of ‘*Ca. Epulopiscium viviparous*.’ Biochemical and other functional analyses need to be conducted to confirm the numerous hypotheses generated from this study.

5.6 REFERENCES

1. Clements KD, Angert ER, Montgomery WL, Choat JH. 2014. Intestinal microbiota in fishes: what's known and what's not. *Molecular ecology* 23:1891-1898.
2. Clements K, Sutton D, Choat J. 1989. Occurrence and characteristics of unusual protistan symbionts from surgeonfishes (Acanthuridae) of the Great Barrier Reef, Australia. *Marine Biology* 102:403-412.
3. Miyake S, Ngugi DK, Stingl U. 2015. Diet strongly influences the gut microbiota of surgeonfishes. *Molecular ecology* 24:656-672.

4. Wang AR, Ran C, Ringø E, Zhou ZG. 2017. Progress in fish gastrointestinal microbiota research. *Reviews in Aquaculture*.
5. Sullam KE, Rubin BE, Dalton CM, Kilham SS, Flecker AS, Russell JA. 2015. Divergence across diet, time and populations rules out parallel evolution in the gut microbiomes of Trinidadian guppies. *The ISME journal* 9:1508-1522.
6. Nayak SK. 2010. Role of gastrointestinal microbiota in fish. *Aquaculture Research* 41:1553-1573.
7. Stephens WZ, Burns AR, Stagaman K, Wong S, Rawls JF, Guillemin K, Bohannan BJ. 2016. The composition of the zebrafish intestinal microbial community varies across development. *The ISME journal* 10:644.
8. Li X, Yan Q, Ringø E, Wu X, He Y, Yang D. 2016. The influence of weight and gender on intestinal bacterial community of wild largemouth bronze gudgeon (*Coreius guichenoti*, 1874). *BMC microbiology* 16:191.
9. Lyons PP, Turnbull JF, Dawson KA, Crumlish M. 2015. Exploring the microbial diversity of the distal intestinal lumen and mucosa of farmed rainbow trout *Oncorhynchus mykiss* (Walbaum) using next generation sequencing (NGS). *Aquaculture Research*.
10. Sullam KE, Essinger SD, Lozupone CA, O'CONNOR MP, Rosen GL, Knight R, Kilham SS, Russell JA. 2012. Environmental and ecological factors that shape the gut bacterial communities of fish: a meta-analysis. *Molecular ecology* 21:3363-3378.
11. Rawls JF, Samuel BS, Gordon JI. 2004. Gnotobiotic zebrafish reveal evolutionarily conserved responses to the gut microbiota. *Proceedings of the National Academy of Sciences of the United States of America* 101:4596-4601.
12. Bates JM, Mittge E, Kuhlman J, Baden KN, Cheesman SE, Guillemin K. 2006. Distinct signals from the microbiota promote different aspects of zebrafish gut differentiation. *Developmental biology* 297:374-386.
13. Cheesman SE, Neal JT, Mittge E, Seredick BM, Guillemin K. 2011. Epithelial cell proliferation in the developing zebrafish intestine is regulated by the Wnt pathway and microbial signaling via Myd88. *Proceedings of the National Academy of Sciences* 108:4570-4577.
14. Pérez T, Balcázar J, Ruiz-Zarzuela I, Halaihel N, Vendrell D, De Blas I, Múzquiz J. 2010. Host-microbiota interactions within the fish intestinal ecosystem. *Mucosal immunology*.
15. Sugita H, Miyajima C, Deguchi Y. 1991. The vitamin B12-producing ability of the intestinal microflora of freshwater fish. *Aquaculture* 92:267-276.
16. Jhaveri P, Papastamatiou YP, German DP. 2015. Digestive enzyme activities in the guts of bonnethead sharks (*Sphyrna tiburo*) provide insight into their digestive strategy and evidence for microbial digestion in their hindguts. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* 189:76-83.
17. Mountfort DO, Campbell J, Clements KD. 2002. Hindgut fermentation in three species of marine herbivorous fish. *Applied and environmental microbiology* 68:1374-1380.
18. Clements K, Choat J. 1995. Fermentation in tropical marine herbivorous fishes. *Physiological Zoology* 68:355-378.
19. Clements KD, Raubenheimer D, Choat JH. 2009. Nutritional ecology of marine herbivorous fishes: ten years on. *Functional Ecology* 23:79-92.

20. Clements KD, Pasch IB, Moran D, Turner SJ. 2007. Clostridia dominate 16S rRNA gene libraries prepared from the hindgut of temperate marine herbivorous fishes. *Marine Biology* 150:1431-1440.
21. Flint JF, Drzymalski D, Montgomery WL, Southam G, Angert ER. 2005. Nocturnal production of endospores in natural populations of *Epulopiscium*-like surgeonfish symbionts. *Journal of bacteriology* 187:7460-7470.
22. Angert ER, Clements KD, Pace NR. 1993. The largest bacterium. *Nature* 362:239-241.
23. Angert ER, Clements KD. 2004. Initiation of intracellular offspring in *Epulopiscium*. *Molecular microbiology* 51:827-835.
24. Angert ER. 2006. The enigmatic cytoarchitecture of *Epulopiscium* spp, p 285-301, *Complex Intracellular Structures in Prokaryotes*. Springer.
25. Fishelson L, Montgomery WL, Myrberg Jr AA. 1985. A unique symbiosis in the gut of tropical herbivorous surgeonfish (*Acanthuridae*: *Teleostei*) from the Red Sea. *Science (Washington)* 229:49-51.
26. Schulz HN, Jørgensen BB. 2001. Big bacteria. *Annual Reviews in Microbiology* 55:105-137.
27. Miller DA, Suen G, Clements KD, Angert ER. 2012. The genomic basis for the evolution of a novel form of cellular reproduction in the bacterium *Epulopiscium*. *BMC genomics* 13:265.
28. Mendell JE, Clements KD, Choat JH, Angert ER. 2008. Extreme polyploidy in a large bacterium. *Proceedings of the National Academy of Sciences* 105:6730-6734.
29. Robinow C, Angert ER. 1998. Nucleoids and coated vesicles of “*Epulopiscium*” spp. *Archives of microbiology* 170:227-235.
30. Tan IS, Ramamurthi KS. 2014. Spore formation in *Bacillus subtilis*. *Environmental microbiology reports* 6:212-225.
31. Angert ER, Losick RM. 1998. Propagation by sporulation in the guinea pig symbiont *Metabacterium polyspora*. *Proceedings of the National Academy of Sciences* 95:10218-10223.
32. Ward RJ, Clements KD, Choat JH, Angert ER. 2009. Cytology of terminally differentiated *Epulopiscium* mother cells. *DNA and cell biology* 28:57-64.
33. Turner S, Pryer KM, Miao VP, Palmer JD. 1999. Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *J Eukaryot Microbiol* 46:327-338.
34. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature methods* 10:563-569.
35. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome research* 27:722-736.
36. Chevreux B, Wetter T, Suhai S. Genome sequence assembly using trace signals and additional sequence information, p 45-56. *In* (ed), Hanover, Germany,
37. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology* 19:455-477.

38. Safonova Y, Bankevich A, Pevzner PA. 2015. dipSPAdes: assembler for highly polymorphic diploid genomes. *Journal of Computational Biology* 22:528-545.
39. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome Research* 27:824-834.
40. Salmela L, Rivals E. 2014. LoRDEC: accurate and efficient long read error correction. *Bioinformatics* 30:3506-3514.
41. Chaisson MJ, Tesler G. 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC bioinformatics* 13:238.
42. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research* 25:1043-1055.
43. Dupont CL, Rusch DB, Yooseph S, Lombardo M-J, Richter RA, Valas R, Novotny M, Yee-Greenbaum J, Selengut JD, Haft DH. 2012. Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *The ISME journal* 6:1186-1199.
44. Gil R, Silva FJ, Peretó J, Moya A. 2004. Determination of the core of a minimal bacterial gene set. *Microbiology and Molecular Biology Reviews* 68:518-537.
45. Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. 2008. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 25:119-120.
46. Brown CT, Olm MR, Thomas BC, Banfield JF. 2016. Measurement of bacterial replication rates in microbial communities. *Nature biotechnology* 34:1256-1263.
47. Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P. 2012. IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic acids research* 40:D115-D122.
48. Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D, Palaniappan K, Szeto E, Pillay M, Chen I-MA, Pati A. 2015. The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v. 4). *Standards in genomic sciences* 10:86.
49. Vallenet D, Belda E, Calteau A, Cruveiller S, Engelen S, Lajus A, Le Fèvre F, Longin C, Mornico D, Roche D. 2012. MicroScope—an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. *Nucleic acids research:gks1194*.
50. Rose PW, Prlić A, Altunkaya A, Bi C, Bradley AR, Christie CH, Costanzo LD, Duarte JM, Dutta S, Feng Z. 2016. The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic acids research:gkw1000*.
51. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic acids research* 40:W445-W451.
52. Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods* 8:785-786.
53. Saier MH, Reddy VS, Tsu BV, Ahmed MS, Li C, Moreno-Hagelsieb G. 2016. The transporter classification database (TCDB): recent advances. *Nucleic Acids Research* 44:D372-D379.
54. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236-1240.

55. Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic acids research* 27:573.
56. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics* 11:119.
57. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy T, Schulz F, Jarett J, Rivers AR, Eloie-Fadrosch EA. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature biotechnology* 35:725.
58. Arakawa K, Tomita M. 2007. The GC skew index: a measure of genomic compositional asymmetry and the degree of replicational selection. *Evolutionary bioinformatics online* 3:159.
59. Miller DA, Choat JH, Clements KD, Angert ER. 2011. The spoIIE homolog of *Epulopiscium* sp. type B is expressed early in intracellular offspring development. *Journal of bacteriology* 193:2642-2646.
60. Kuwana R, Takamatsu H. 2013. The GerW protein is essential for L-alanine-stimulated germination of *Bacillus subtilis* spores. *The Journal of Biochemistry* 154:409-417.
61. Kodama T, Takamatsu H, Asai K, Kobayashi K, Ogasawara N, Watabe K. 1999. The *Bacillus subtilis* yaaH gene is transcribed by SigE RNA polymerase during sporulation, and its product is involved in germination of spores. *Journal of bacteriology* 181:4584-4591.
62. Eichenberger P, Jensen ST, Conlon EM, Van Ooij C, Silvaggi J, Gonzalez-Pastor J-E, Fujita M, Ben-Yehuda S, Stragier P, Liu JS. 2003. The σ E regulon and the identification of additional sporulation genes in *Bacillus subtilis*. *Journal of molecular biology* 327:945-972.
63. Kuwana R, Kasahara Y, Fujibayashi M, Takamatsu H, Ogasawara N, Watabe K. 2002. Proteomics characterization of novel spore proteins of *Bacillus subtilis*. *Microbiology* 148:3971-3982.
64. Eichenberger P, Fujita M, Jensen ST, Conlon EM, Rudner DZ, Wang ST, Ferguson C, Haga K, Sato T, Liu JS. 2004. The program of gene transcription for a single differentiating cell type during sporulation in *Bacillus subtilis*. *PLoS biology* 2:e328.
65. Al-Hinai MA, Jones SW, Papoutsakis ET. 2015. The *Clostridium* sporulation programs: diversity and preservation of endospore differentiation. *Microbiology and Molecular Biology Reviews* 79:19-37.
66. Steiner E, Dago AE, Young DI, Heap JT, Minton NP, Hoch JA, Young M. 2011. Multiple orphan histidine kinases interact directly with Spo0A to control the initiation of endospore formation in *Clostridium acetobutylicum*. *Molecular microbiology* 80:641-654.
67. Lane N, Martin W. 2010. The energetics of genome complexity. *Nature* 467:929-934.
68. Reva O, Tümmler B. 2008. Think big—giant genes in bacteria. *Environmental microbiology* 10:768-777.
69. López-Pérez M, Gonzaga A, Martin-Cuadrado A-B, Onyshchenko O, Ghavidel A, Ghai R, Rodríguez-Valera F. 2012. Genomes of surface isolates of *Alteromonas macleodii*: the life of a widespread marine opportunistic copiotroph. *Scientific reports* 2.

70. Scanlan DJ, Ostrowski M, Mazard S, Dufresne A, Garczarek L, Hess WR, Post AF, Hagemann M, Paulsen I, Partensky F. 2009. Ecological genomics of marine picocyanobacteria. *Microbiology and Molecular Biology Reviews* 73:249-299.
71. Strom SL, Brahamsha B, Fredrickson KA, Apple JK, Rodríguez AG. 2012. A giant cell surface protein in *Synechococcus* WH8102 inhibits feeding by a dinoflagellate predator. *Environmental microbiology* 14:807-816.
72. Grim NJ. 2006. Food vacuole contents in the ciliate, *Balantidium jocularum* (Balantidiidae), a symbiont in the intestine of the surgeonfish, *Naso tonganus* (Acanthuridae). *Journal of Eukaryotic Microbiology* 53:269-274.
73. Vogl K, Wenter R, Dreßen M, Schlickenrieder M, Plösch M, Eichacker L, Overmann J. 2008. Identification and analysis of four candidate symbiosis genes from ‘*Chlorochromatium aggregatum*’, a highly developed bacterial symbiosis. *Environmental microbiology* 10:2842-2856.
74. Wagner C, Polke M, Gerlach RG, Linke D, Stierhof YD, Schwarz H, Hensel M. 2011. Functional dissection of SiiE, a giant non-fimbrial adhesin of *Salmonella enterica*. *Cellular microbiology* 13:1286-1301.
75. Freudl R. 2013. Leaving home ain't easy: protein export systems in Gram-positive bacteria. *Research in microbiology* 164:664-674.
76. Ates LS, Houben EN, Bitter W. 2016. Type VII secretion: a highly versatile secretion system. *Microbiology spectrum* 4.
77. Fyans JK, Bignell D, Loria R, Toth I, Palmer T. 2013. The ESX/type VII secretion system modulates development, but not virulence, of the plant pathogen *Streptomyces scabies*. *Molecular plant pathology* 14:119-130.
78. Huppert LA, Ramsdell TL, Chase MR, Sarracino DA, Fortune SM, Burton BM. 2014. The ESX system in *Bacillus subtilis* mediates protein secretion. *PloS one* 9:e96267.
79. Bottai D, Gröschel MI, Brosch R. 2015. Type VII secretion systems in gram-positive bacteria.
80. Apel D, Surette MG. 2008. Bringing order to a complex molecular machine: the assembly of the bacterial flagella. *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1778:1851-1858.
81. Mukherjee S, Kearns DB. 2014. The structure and regulation of flagella in *Bacillus subtilis*. *Annual review of genetics* 48:319-340.
82. Altegoer F, Bange G. 2015. Undiscovered regions on the molecular landscape of flagellar assembly. *Current opinion in microbiology* 28:98-105.
83. Liu R, Ochman H. 2007. Stepwise formation of the bacterial flagellar system. *Proceedings of the National Academy of Sciences* 104:7116-7121.
84. Ito M, Hicks DB, Henkin TM, Guffanti AA, Powers BD, Zvi L, Uematsu K, Krulwich TA. 2004. MotPS is the stator-force generator for motility of alkaliphilic *Bacillus*, and its homologue is a second functional Mot in *Bacillus subtilis*. *Molecular microbiology* 53:1035-1049.
85. McQuiston JR, Fields PI, Tauxe RV, Logsdon JM. 2008. Do *Salmonella* carry spare tyres? *Trends in microbiology* 16:142-148.
86. Rossez Y, Wolfson EB, Holmes A, Gally DL, Holden NJ. 2015. Bacterial flagella: twist and stick, or dodge across the kingdoms. *PLoS Pathog* 11:e1004483.
87. Rao CV, Glekas GD, Ordal GW. 2008. The three adaptation systems of *Bacillus subtilis* chemotaxis. *Trends in microbiology* 16:480-487.

88. Ud-Din AIMS, Roujeinikova A. 2017. Methyl-accepting chemotaxis proteins: a core sensing element in prokaryotes and archaea. *Cellular and Molecular Life Sciences*:1-11.
89. Bonfini A, Liu X, Buchon N. 2016. From pathogens to microbiota: How *Drosophila* intestinal stem cells react to gut microbes. *Developmental & Comparative Immunology* 64:22-38.
90. Ruiz N. 2015. Lipid flippases for bacterial peptidoglycan biosynthesis. *Lipid insights* 8:21.
91. Sham L-T, Butler EK, Lebar MD, Kahne D, Bernhardt TG, Ruiz N. 2014. MurJ is the flippase of lipid-linked precursors for peptidoglycan biogenesis. *Science* 345:220-222.
92. Scheffers D-J, Pinho MG. 2005. Bacterial cell wall synthesis: new insights from localization studies. *Microbiology and Molecular Biology Reviews* 69:585-607.
93. Brown S, Santa Maria Jr JP, Walker S. 2013. Wall teichoic acids of gram-positive bacteria. *Annual review of microbiology* 67:313-336.
94. Reichmann NT, Gründling A. 2011. Location, synthesis and function of glycolipids and polyglycerolphosphate lipoteichoic acid in Gram-positive bacteria of the phylum Firmicutes. *FEMS microbiology letters* 319:97-105.
95. Percy MG, Gründling A. 2014. Lipoteichoic acid synthesis and function in gram-positive bacteria. *Annual review of microbiology* 68:81-100.
96. Candela T, Fouet A. 2006. Poly-gamma-glutamate in bacteria. *Molecular microbiology* 60:1091-1098.
97. Dürre P. 2014. Physiology and sporulation in *Clostridium*. *Microbiology spectrum* 2.
98. Crown SB, Indurthi DC, Ahn WS, Choi J, Papoutsakis ET, Antoniewicz MR. 2011. Resolving the TCA cycle and pentose-phosphate pathway of *Clostridium acetobutylicum* ATCC 824: Isotopomer analysis, in vitro activities and expression analysis. *Biotechnology journal* 6:300-305.
99. Sridhar J, Eiteman MA, Wiegel JW. 2000. Elucidation of enzymes in fermentation pathways used by *Clostridium thermosuccinogenes* growing on inulin. *Applied and environmental microbiology* 66:246-251.
100. Buckel W. 2001. Sodium ion-translocating decarboxylases. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1505:15-27.
101. Woehlke G, Dimroth P. 1994. Anaerobic growth of *Salmonella typhimurium* on l (+)- and d (-)-tartrate involves an oxaloacetate decarboxylase Na⁺ pump. *Archives of microbiology* 162:233-237.
102. Chen Y-T, Liao T-L, Wu K-M, Lauderdale T-L, Yan J-J, Huang I-W, Lu M-C, Lai Y-C, Liu Y-M, Shu H-Y. 2009. Genomic diversity of citrate fermentation in *Klebsiella pneumoniae*. *BMC microbiology* 9:168.
103. Bott M, Dimroth P. 1994. *Klebsiella pneumoniae* genes for citrate lyase and citrate lyase ligase: localization, sequencing, and expression. *Molecular microbiology* 14:347-356.
104. Warner JB, Lolkema JS. 2002. Growth of *Bacillus subtilis* on citrate and isocitrate is supported by the Mg²⁺-citrate transporter CitM. *Microbiology* 148:3405-3412.
105. Dimroth P, Schink B. 1998. Energy conservation in the decarboxylation of dicarboxylic acids by fermenting bacteria. *Archives of microbiology* 170:69-77.
106. Hilpert W, Schink B, Dimroth P. 1984. Life by a new decarboxylation-dependent energy conservation mechanism with Na⁺ as coupling ion. *The EMBO journal* 3:1665.

107. Schulz S, Iglesias-Cans M, Krah A, Yildiz Ö, Leone V, Matthies D, Cook GM, Faraldo-Gómez JD, Meier T. 2013. A new type of Na⁺-driven ATP synthase membrane rotor with a two-carboxylate ion-coupling motif. *PLoS Biol* 11:e1001596.
108. Meier T, Polzer P, Diederichs K, Welte W, Dimroth P. 2005. Structure of the rotor ring of F-Type Na⁺-ATPase from *Ilyobacter tartaricus*. *Science* 308:659-662.
109. Barbeyron T, Thomas F, Barbe V, Teeling H, Schenowitz C, Dossat C, Goesmann A, Leblanc C, Oliver Glöckner F, Czjzek M. 2016. Habitat and taxon as driving forces of carbohydrate catabolism in marine heterotrophic bacteria: example of the model algae-associated bacterium *Zobellia galactanivorans* DsijT. *Environmental microbiology* 18:4610-4627.
110. Mann AJ, Hahnke RL, Huang S, Werner J, Xing P, Barbeyron T, Huettel B, Stüber K, Reinhardt R, Harder J. 2013. The genome of the alga-associated marine flavobacterium *Formosa agariphila* KMM 3901T reveals a broad potential for degradation of algal polysaccharides. *Applied and environmental microbiology* 79:6813-6822.
111. Choat J, Clements K, Robbins W. 2002. The trophic status of herbivorous fishes on coral reefs. *Marine Biology* 140:613-623.
112. Chi W-J, Chang Y-K, Hong S-K. 2012. Agar degradation by microorganisms and agar-degrading enzymes. *Applied microbiology and biotechnology* 94:917-930.
113. Fu XT, Kim SM. 2010. Agarase: review of major sources, categories, purification method, enzyme characteristics and applications. *Marine drugs* 8:200-218.
114. Michel G, Nyval-Collen P, Barbeyron T, Czjzek M, Helbert W. 2006. Bioconversion of red seaweed galactans: a focus on bacterial agarases and carrageenases. *Applied Microbiology and Biotechnology* 71:23-33.
115. Hehemann J-H, Correc G, Thomas F, Bernard T, Barbeyron T, Jam M, Helbert W, Michel G, Czjzek M. 2012. Biochemical and structural characterization of the complex agarolytic enzyme system from the marine bacterium *Zobellia galactanivorans*. *Journal of Biological Chemistry* 287:30571-30584.
116. Hutcheson SW, Zhang H, Suvorov M. 2011. Carbohydrase systems of *Saccharophagus degradans* degrading marine complex polysaccharides. *Marine drugs* 9:645-665.
117. Hehemann J-H, Boraston AB, Czjzek M. 2014. A sweet new wave: structures and mechanisms of enzymes that digest polysaccharides from marine algae. *Current opinion in structural biology* 28:77-86.
118. Ha SC, Lee S, Lee J, Kim HT, Ko H-J, Kim KH, Choi I-G. 2011. Crystal structure of a key enzyme in the agarolytic pathway, α -neoagarobiose hydrolase from *Saccharophagus degradans* 2–40. *Biochemical and biophysical research communications* 412:238-244.
119. Lee SB, Cho SJ, Kim JA, Lee SY, Kim SM, Lim HS. 2014. Metabolic pathway of 3, 6-anhydro-L-galactose in agar-degrading microorganisms. *Biotechnology and bioprocess engineering* 19:866-878.
120. Hobbs ME, Williams HJ, Hillerich B, Almo SC, Raushel FM. 2014. l-Galactose metabolism in *Bacteroides vulgatus* from the human gut microbiota. *Biochemistry* 53:4661-4670.
121. Guibet M, Colin S, Barbeyron T, Genicot S, Kloareg B, Michel G, Helbert W. 2007. Degradation of λ -carrageenan by *Pseudoalteromonas carrageenovora* λ -carrageenase: a new family of glycoside hydrolases unrelated to κ - and ι -carrageenases. *Biochemical Journal* 404:105-114.

122. Chauhan PS, Saxena A. 2016. Bacterial carrageenases: an overview of production and biotechnological applications. 3 Biotech 6:1-18.
123. Lee SB, Kim JA, Lim HS. 2016. Metabolic pathway of 3, 6-anhydro-D-galactose in carrageenan-degrading microorganisms. Applied microbiology and biotechnology 100:4109-4121.
124. Xing P, Hahnke RL, Unfried F, Markert S, Huang S, Barbeyron T, Harder J, Becher D, Schweder T, Glöckner FO. 2015. Niches of two polysaccharide-degrading *Polaribacter* isolates from the North Sea during a spring diatom bloom. The ISME journal 9:1410.
125. Lechat H, Amat M, Mazoyer J, Buléon A, Lahaye M. 2000. Structure and distribution of glucomannan and sulfated glucan in the cell walls of the red alga *Kappaphycus alvarezii* (Gigartinales, Rhodophyta). Journal of Phycology 36:891-902.
126. Turvey J, Williams E. 1970. The structures of some xylans from red algae. Phytochemistry 9:2383-2388.
127. Cartmell A, Topakas E, Ducros VM, Suits MD, Davies GJ, Gilbert HJ. 2008. The *Cellvibrio japonicus* mannanase CjMan26C displays a unique exo-mode of action that is conferred by subtle changes to the distal region of the active site. Journal of Biological Chemistry 283:34403-34413.
128. Yamagaki T, Maeda M, Kanazawa K, Ishizuka Y, Nakanishi H. 1997. Structural clarification of *Caulerpa* cell wall β -1, 3-xylan by NMR spectroscopy. Bioscience, biotechnology, and biochemistry 61:1077-1080.
129. Domozych DS, Ciancia M, Fangel JU, Mikkelsen MD, Ulvskov P, Willats WG. 2012. The Cell Walls of Green Algae: A Journey through Evolution and Diversity. Front Plant Sci 3:82.
130. Estevez JM, Leonardi PI, Alberghina JS. 2008. Cell wall carbohydrate epitopes in the green alga *Oedogonium bharuchae* F. *minor* (Oedogoniales, Chlorophyta)¹. Journal of phycology 44:1257-1268.
131. Ciancia M, Quintana I, Vizcargüénaga MI, Kasulin L, de Dios A, Estevez JM, Cerezo AS. 2007. Polysaccharides from the green seaweeds *Codium fragile* and *C. vermilara* with controversial effects on hemostasis. International Journal of Biological Macromolecules 41:641-649.
132. Lahaye M, Robic A. 2007. Structure and functional properties of ulvan, a polysaccharide from green seaweeds. Biomacromolecules 8:1765-1774.
133. Abbott DW, Hrynuik S, Boraston AB. 2007. Identification and characterization of a novel periplasmic polygalacturonic acid binding protein from *Yersinia enterocolitica*. Journal of molecular biology 367:1023-1033.
134. Ichinose H, Araki Y, Michikawa M, Harazono K, Yaoi K, Karita S, Kaneko S. 2012. Characterization of an endo-processive-type xyloglucanase having a β -1, 4-glucan-binding module and an endo-type xyloglucanase from *Streptomyces avermitilis*. Applied and environmental microbiology 78:7939-7945.
135. Attia M, Stepper J, Davies GJ, Brumer H. 2016. Functional and structural characterization of a potent GH74 endo-xyloglucanase from the soil saprophyte *Cellvibrio japonicus* unravels the first step of xyloglucan degradation. FEBS Journal 283:1701-1719.
136. Fry SC, Mohler KE, Nesselrode BH, Franková L. 2008. Mixed-linkage β -glucan: xyloglucan endotransglucosylase, a novel wall-remodelling enzyme from *Equisetum* (horsetails) and charophytic algae. The Plant Journal 55:240-252.

137. Enkhbaatar B, Temuujin U, Lim J-H, Chi W-J, Chang Y-K, Hong S-K. 2012. Identification and characterization of a xyloglucan-specific family 74 glycosyl hydrolase from *Streptomyces coelicolor* A3 (2). *Applied and environmental microbiology* 78:607-611.
138. Hawthorne DB, Sawyer WH, Grant BR. 1979. The structure of the low molecular-weight glucans isolated from the siphonous green alga *Caulerpa simpliciuscula*. *Carbohydrate Research* 77:157-167.
139. Raven JA, Beardall J. 2003. Carbohydrate metabolism and respiration in algae, p 205-224, *Photosynthesis in algae*. Springer.
140. Percival E. 1979. The polysaccharides of green, red and brown seaweeds: their basic structure, biosynthesis and function. *British Phycological Journal* 14:103-117.
141. Domozych DS. 2011. Algal cell walls. eLS.
142. Ji S-Q, Wang B, Lu M, Li F-L. 2016. *Defluviitalea phaphyphila* sp. nov., a Novel Thermophilic Bacterium That Degrades Brown Algae. *Applied and environmental microbiology* 82:868-877.
143. Kornberg HL, Lambourne LT. 1992. Role of the phosphoenolpyruvate-dependent fructose phosphotransferase system in the utilization of mannose by *Escherichia coli*. *Proceedings of the Royal Society of London B: Biological Sciences* 250:51-55.
144. Conejo MS, Thompson SM, Miller BG. 2010. Evolutionary bases of carbohydrate recognition and substrate discrimination in the ROK protein family. *Journal of molecular evolution* 70:545-556.
145. Schneider E. 2001. ABC transporters catalyzing carbohydrate uptake. *Research in microbiology* 152:303-310.
146. Busi MV, Barchiesi J, Martín M, Gomez-Casati DF. 2014. Starch metabolism in green algae. *Starch-Stärke* 66:28-40.
147. Michel G, Tonon T, Scornet D, Cock JM, Kloareg B. 2010. Central and storage carbon metabolism of the brown alga *Ectocarpus siliculosus*: insights into the origin and evolution of storage carbohydrates in Eukaryotes. *New Phytologist* 188:67-81.
148. Barbier G, Oesterhelt C, Larson MD, Hølgren RG, Wilkerson C, Garavito RM, Benning C, Weber AP. 2005. Comparative genomics of two closely related unicellular thermoacidophilic red algae, *Galdieria sulphuraria* and *Cyanidioschyzon merolae*, reveals the molecular basis of the metabolic flexibility of *Galdieria sulphuraria* and significant differences in carbohydrate metabolism of both algae. *Plant physiology* 137:460-474.
149. Holzinger A, Karsten U. 2013. Desiccation stress and tolerance in green algae: consequences for ultrastructure, physiological and molecular mechanisms. *Frontiers in plant science* 4:327.
150. Salerno GL, Pontis HG. 1989. Raffinose synthesis in *Chlorella vulgaris* cultures after a cold shock. *Plant physiology* 89:648-651.
151. Anastyuk SD, Shevchenko NM, Nazarenko EL, Dmitrenok PS, Zvyagintseva TN. 2009. Structural analysis of a fucoidan from the brown alga *Fucus evanescens* by MALDI-TOF and tandem ESI mass spectrometry. *Carbohydrate research* 344:779-787.
152. Utz CB, Nguyen AB, Smalley DJ, Anderson AB, Conway T. 2004. GntP is the *Escherichia coli* fructuronic acid transporter and belongs to the UxuR regulon. *Journal of bacteriology* 186:7690-7696.

153. Karsten U, King RJ, Kirst GO. 1990. The distribution of D-sorbitol and D-dulcitol in the red algal genera *Bostrychia* and *Stictosiphonia* (Rhodomelaceae, Rhodophyta)—a re-evaluation. *British Phycological Journal* 25:363-366.
154. Iwamoto K, Shiraiwa Y. 2005. Salt-regulated mannitol metabolism in algae. *Marine Biotechnology* 7:407-415.
155. Kremer BP. 1980. Taxonomic implications of algal photoassimilate patterns. *British Phycological Journal* 15:399-409.
156. Ferla MP, Patrick WM. 2014. Bacterial methionine biosynthesis. *Microbiology* 160:1571-1584.
157. Rodionov DA, Hebbeln P, Eudes A, Ter Beek J, Rodionova IA, Erkens GB, Slotboom DJ, Gelfand MS, Osterman AL, Hanson AD. 2009. A novel class of modular transporters for vitamins in prokaryotes. *Journal of bacteriology* 191:42-51.
158. Liu Y, White RH, Whitman WB. 2010. Methanococci use the diaminopimelate aminotransferase (DapL) pathway for lysine biosynthesis. *Journal of bacteriology* 192:3304-3310.
159. Hudson AO, Singh BK, Leustek T, Gilvarg C. 2006. An LL-diaminopimelate aminotransferase defines a novel variant of the lysine biosynthesis pathway in plants. *Plant Physiology* 140:292-301.
160. Triassi AJ, Wheatley MS, Savka MA, Gan HM, Dobson RC, Hudson AO. 2014. L, L-diaminopimelate aminotransferase (DapL): a putative target for the development of narrow-spectrum antibacterial compounds. *Frontiers in microbiology* 5.
161. Hudson AO, Gilvarg C, Leustek T. 2008. Biochemical and phylogenetic characterization of a novel diaminopimelate biosynthesis pathway in prokaryotes identifies a diverged form of LL-diaminopimelate aminotransferase. *Journal of bacteriology* 190:3256-3263.
162. Basavanna S, Khandavilli S, Yuste J, Cohen JM, Hosie AH, Webb AJ, Thomas GH, Brown JS. 2009. Screening of *Streptococcus pneumoniae* ABC transporter mutants demonstrates that LivJHMGF, a branched-chain amino acid ABC transporter, is necessary for disease pathogenesis. *Infection and immunity* 77:3412-3423.
163. Wang W, Cho HS, Kim R, Jancarik J, Yokota H, Nguyen HH, Grigoriev IV, Wemmer DE, Kim S-H. 2002. Structural characterization of the reaction pathway in phosphoserine phosphatase: crystallographic “snapshots” of intermediate states. *Journal of molecular biology* 319:421-431.
164. Burguière P, Auger S, Hullo M-F, Danchin A, Martin-Verstraete I. 2004. Three different systems participate in L-cystine uptake in *Bacillus subtilis*. *Journal of bacteriology* 186:4875-4884.
165. Reig N, del Rio C, Casagrande F, Ratera M, Gelpí JL, Torrents D, Henderson PJ, Xie H, Baldwin SA, Zorzano A. 2007. Functional and Structural Characterization of the First Prokaryotic Member of the L-Amino Acid Transporter (LAT) Family A MODEL FOR APC TRANSPORTERS. *Journal of biological chemistry* 282:13270-13281.
166. Rodríguez-Banqueri A, Errasti-Murugarren E, Bartoccioni P, Kowalczyk L, Perálvarez-Marín A, Palacín M, Vázquez-Ibar JL. 2016. Stabilization of a prokaryotic LAT transporter by random mutagenesis. *The Journal of general physiology* 147:353-368.
167. Xu Y, Labedan B, Glansdorff N. 2007. Surprising arginine biosynthesis: a reappraisal of the enzymology and evolution of the pathway in microorganisms. *Microbiology and molecular biology reviews* 71:36-47.

168. Johnson DA, Tetu SG, Phillippy K, Chen J, Ren Q, Paulsen IT. 2008. High-throughput phenotypic characterization of *Pseudomonas aeruginosa* membrane transport genes. *PLoS Genet* 4:e1000211.
169. Sekowska A, Robin S, Daudin J-J, Hénaut A, Danchin A. 2001. Extracting biological information from DNA arrays: an unexpected link between arginine and methionine metabolism in *Bacillus subtilis*. *Genome biology* 2:research0019. 1.
170. Ratzkin B, Roth J. 1978. Cluster of genes controlling proline degradation in *Salmonella typhimurium*. *Journal of bacteriology* 133:744-754.
171. Abrahamson JL, Baker LG, Stephenson JT, Wood JM. 1983. Proline Dehydrogenase from *Escherichia coli* K12. *European Journal of Biochemistry* 134:77-82.
172. Neumann-Schaal M, Hofmann JD, Will SE, Schomburg D. 2015. Time-resolved amino acid uptake of *Clostridium difficile* 630 Δ erm and concomitant fermentation product and toxin formation. *BMC microbiology* 15:281.
173. Jackson S, Calos M, Myers A, Self WT. 2006. Analysis of proline reduction in the nosocomial pathogen *Clostridium difficile*. *Journal of bacteriology* 188:8487-8495.
174. Andreesen JR, Wagner M, Sonntag D, Kohlstock M, Harms C, Gursinsky T, Jäge J, Parther T, Kabisch U, Gräntzdörffer A. 1999. Various functions of selenols and thiols in anaerobic Gram-positive, amino acids-utilizing bacteria. *Biofactors* 10:263-270.
175. Simanshu DK, Chittori S, Savithri H, Murthy MR. 2007. Structure and function of enzymes involved in the anaerobic degradation of L-threonine to propionate. *Journal of biosciences* 32:1195-1206.
176. Fong YH, Wong HC, Yuen MH, Lau PH, Chen YW, Wong K-B. 2013. Structure of UreG/UreF/UreH complex reveals how urease accessory proteins facilitate maturation of *Helicobacter pylori* urease. *PLoS Biol* 11:e1001678.
177. Mobley H, Hausinger R. 1989. Microbial ureases: significance, regulation, and molecular characterization. *Microbiological reviews* 53:85-108.
178. Matsumoto M, Kibe R, Ooga T, Aiba Y, Kurihara S, Sawaki E, Koga Y, Benno Y. 2012. Impact of intestinal microbiota on intestinal luminal metabolome. *Scientific reports* 2:233.
179. Ip AY, Chew SF. 2010. Ammonia production, excretion, toxicity, and defense in fish: a review. *Frontiers in Physiology* 1:134.
180. Barimo JF, Walsh PJ. 2006. Use of urea as a chemosensory cloaking molecule by a bony fish. *Journal of experimental biology* 209:4254-4261.
181. Bucking C, LeMoine CM, Craig PM, Walsh PJ. 2013. Nitrogen metabolism of the intestine during digestion in a teleost fish, the plainfin midshipman (*Porichthys notatus*). *Journal of Experimental Biology* 216:2821-2832.
182. Fujinaga K, Taniguchi Y, Sun Y, Katayama S, Minami J, Matsushita O, Okabe A. 1999. Analysis of genes involved in nitrate reduction in *Clostridium perfringens*. *Microbiology* 145:3377-3387.
183. Moreno-Vivián C, Cabello P, Martínez-Luque M, Blasco R, Castillo F. 1999. Prokaryotic nitrate reduction: molecular properties and functional distinction among bacterial nitrate reductases. *Journal of bacteriology* 181:6573-6584.
184. Seki S, Satoh M, Seki Y, Yamamoto I, Kondo H. 1989. Nitrate reduction by pyruvate in *Clostridium perfringens*. *The Journal of General and Applied Microbiology* 35:107-112.
185. Hasan S, Hall J. 1975. The physiological function of nitrate reduction in *Clostridium perfringens*. *Microbiology* 87:120-128.

186. Ishimoto M, Umeyama M, Chiba S. 1974. Alteration of fermentation products from butyrate to acetate by nitrate reduction in *Clostridium perfringens*. *Zeitschrift für allgemeine Mikrobiologie* 14:115-121.
187. Jia W, Tovell N, Clegg S, Trimmer M, Cole J. 2009. A single channel for nitrate uptake, nitrite export and nitrite uptake by *Escherichia coli* NarU and a role for NirC in nitrite export and uptake. *Biochemical Journal* 417:297-307.
188. Strohmeier M, Raschle T, Mazurkiewicz J, Rippe K, Sinning I, Fitzpatrick TB, Tews I. 2006. Structure of a bacterial pyridoxal 5'-phosphate synthase complex. *Proceedings of the National Academy of Sciences* 103:19284-19289.
189. Koyama M, Katayama S, Kaji M, Taniguchi Y, Matsushita O, Minami J, Morita S, Okabe A. 1999. A *Clostridium perfringens* hem Gene Cluster Contains a *cysGB* Homologue That Is Involved in Cobalamin Biosynthesis. *Microbiology and immunology* 43:947-957.
190. Campbell GR, Taga ME, Mistry K, Lloret J, Anderson PJ, Roth JR, Walker GC. 2006. *Sinorhizobium meliloti* bluB is necessary for production of 5, 6-dimethylbenzimidazole, the lower ligand of B12. *Proceedings of the National Academy of Sciences of the United States of America* 103:4634-4639.
191. Hazra AB, Han AW, Mehta AP, Mok KC, Osadchiy V, Begley TP, Taga ME. 2015. Anaerobic biosynthesis of the lower ligand of vitamin B12. *Proceedings of the National Academy of Sciences* 112:10792-10797.
192. Torres AC, Vannini V, Bonacina J, Font G, Saavedra L, Taranto MP. 2016. Cobalamin production by *Lactobacillus coryniformis*: biochemical identification of the synthesized corrinoid and genomic analysis of the biosynthetic cluster. *BMC microbiology* 16:240.
193. Erkens GB, Majsnerowska M, ter Beek J, Slotboom DJ. 2012. Energy coupling factor-type ABC transporters for vitamin uptake in prokaryotes. *Biochemistry* 51:4390-4396.
194. Karpowich NK, Song JM, Cocco N, Wang D-N. 2015. ATP binding drives substrate capture in an ECF transporter by a release-and-catch mechanism. *Nature structural & molecular biology* 22:565-571.
195. Jurgenson CT, Begley TP, Ealick SE. 2009. The structural and biochemical foundations of thiamin biosynthesis. *Annu Rev Biochem* 78:569-603.
196. Hasnain G, Roje S, Sa N, Zallot R, Ziemak MJ, de Crécy-Lagard V, Gregory JF, Hanson AD. 2016. Bacterial and plant HAD enzymes catalyse a missing phosphatase step in thiamin diphosphate biosynthesis. *Biochemical Journal* 473:157-166.
197. Anderson LN, Koech PK, Plymale AE, Landorf EV, Konopka A, Collart FR, Lipton MS, Romine MF, Wright AT. 2015. Live cell discovery of microbial vitamin transport and enzyme-cofactor interactions. *ACS chemical biology* 11:345-354.
198. Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS. 2002. Comparative genomics of thiamin biosynthesis in procaryotes New genes and regulatory mechanisms. *Journal of Biological chemistry* 277:48949-48959.
199. Jenkins AH, Schyns G, Potot S, Sun G, Begley TP. 2007. A new thiamin salvage pathway. *Nature chemical biology* 3:492-497.
200. Rossi M, Amaretti A, Raimondi S. 2011. Folate production by probiotic bacteria. *Nutrients* 3:118-134.
201. Carter EL, Jager L, Gardner L, Hall CC, Willis S, Green JM. 2007. *Escherichia coli* *abg* genes enable uptake and cleavage of the folate catabolite p-aminobenzoyl-glutamate. *Journal of bacteriology* 189:3329-3334.

202. Delmar JA, Yu EW. 2016. The AbgT family: a novel class of antimetabolite transporters. *Protein Science* 25:322-337.
203. Kurnasov O, Goral V, Colabroy K, Gerdes S, Anantha S, Osterman A, Begley TP. 2003. NAD biosynthesis: identification of the tryptophan to quinolinate pathway in bacteria. *Chemistry & biology* 10:1195-1204.
204. Johnson MD, Echlin H, Dao TH, Rosch JW. 2015. Characterization of NAD salvage pathways and their role in virulence in *Streptococcus pneumoniae*. *Microbiology* 161:2127-2136.
205. Pollack JD, Myers MA, Dandekar T, Herrmann R. 2002. Suspected utility of enzymes with multiple activities in the small genome *Mycoplasma* species: the replacement of the missing "household" nucleoside diphosphate kinase gene and activity by glycolytic kinases. *Omics: a journal of integrative biology* 6:247-258.
206. Kuroda A, Kornberg A. 1997. Polyphosphate kinase as a nucleoside diphosphate kinase in *Escherichia coli* and *Pseudomonas aeruginosa*. *Proceedings of the National Academy of Sciences* 94:439-442.
207. Angert ER. 2012. DNA replication and genomic architecture of very large bacteria. *Annual review of microbiology* 66:197-212.
208. Salman V, Amann R, Girnth A-C, Polerecky L, Bailey JV, Høgslund S, Jessen G, Pantoja S, Schulz-Vogt HN. 2011. A single-cell sequencing approach to the classification of large, vacuolated sulfur bacteria. *Systematic and Applied Microbiology* 34:243-259.
209. Winkel M, Salman-Carvalho V, Woyke T, Richter M, Schulz-Vogt HN, Flood BE, Bailey JV, Mußmann M. 2016. Single-cell sequencing of *Thiomargarita* reveals genomic flexibility for adaptation to dynamic redox conditions. *Frontiers in microbiology* 7.
210. Flood BE, Fliss P, Jones DS, Dick GJ, Jain S, Kaster A-K, Winkel M, Mußmann M, Bailey J. 2016. Single-cell (meta-) genomics of a dimorphic *Candidatus Thiomargarita nelsonii* reveals genomic plasticity. *Frontiers in microbiology* 7.
211. Rodriguez Ayala F, Bauman C, Bartolini M, Saball E, Salvarrey M, Leñini C, Cogliati S, Strauch M, Grau R. 2017. Transcriptional regulation of adhesive properties of *Bacillus subtilis* to extracellular matrix proteins through the fibronectin-binding protein YloA. *Molecular Microbiology*.
212. Timinskas K, Balvočiūtė M, Timinskas A, Venclovas Č. 2013. Comprehensive analysis of DNA polymerase III α subunits and their homologs in bacterial genomes. *Nucleic acids research* 42:1393-1413.
213. Sanders GM, Dallmann HG, McHenry CS. 2010. Reconstitution of the *B. subtilis* replisome with 13 proteins including two distinct replicases. *Molecular cell* 37:273-281.
214. Asplund-Samuelsson J, Bergman B, Larsson J. 2012. Prokaryotic caspase homologs: phylogenetic patterns and functional characteristics reveal considerable diversity. *PLoS One* 7:e49888.
215. Flint HJ, Bayer EA, Rincon MT, Lamed R, White BA. 2008. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nature Reviews Microbiology* 6:121-131.

APPENDIX A

INVESTIGATIONS INTO THE ROLE OF THIAMINASE I AS A PATHOGENICTY AND COMPETITION FACTOR

A.1 INTRODUCTION

Thiamin is essential as it plays crucial roles in carbon transformations within a cell. It serves as a cofactor for enzymes necessary for energy production and the biosynthesis of important precursors for amino acids, nucleotides, and vitamins (1). It can be synthesized by many species of bacteria, archaea, fungi, and plants, but it cannot be synthesized by animals. Thiamin is composed of two moieties, a pyrimidine moiety, hydroxymethyl pyrimidine (HMP), which in bacteria, is derived from 5-aminoimidazole ribotide (AIR), an intermediate in purine biosynthesis, and a thiazole moiety (THZ), which, in bacteria, is derived from pyruvate, glyceraldehyde-3-phosphate, glycine or tyrosine (depending on the species), and a sulfur from cysteine (2). Each moiety is synthesized separately; the HMP moiety is synthesized in one step via the HMP synthase, ThiC, in a complex chemical rearrangement, to form the phosphorylated HMP-P, while THZ is synthesized in multiple steps, with all the precursors being combined to form the phosphorylated THZ-P by the thiazole synthase, ThiG (2). HMP-P is then phosphorylated again by the HMP kinase ThiD to generate HMP-PP, which is then combined with THZ-P by the thiamin phosphate synthase, ThiE, to form thiamin monophosphate (TMP). TMP is either phosphorylated again by ThiL to form the active cofactor TPP, or dephosphorylated to thiamin, and then pyrophosphorylated by thiamin pyrophosphokinase (2). Although many species of bacteria are able to synthesize thiamin, multiple thiamin-auxotrophs occur in nature. These auxotrophs are able to offset their lack of biosynthetic capabilities by the import of thiamin through specific transporters for the uptake of it, and or its phosphorylated forms (3, 4). Other bacteria are able to quench their auxotrophy through salvage pathways (5, 6),

in which thiamin precursors are imported (7-10), phosphorylated, and then combined to form the active cofactor. Thiaminases have been implicated in playing roles in thiamin salvage.

Thiaminases are enzymes that degrade thiamin to its constituent moieties and are categorized into 2, non-homologous classes. Thiaminase II (TenA) enzymes are widespread, cytoplasmic enzymes found in bacteria (11) and archaea (12), and are also found in plants (13) and fungi (14). They catalyze the base substitution of base-degraded thiamin via a nucleophilic attack using water as a nucleophile. This replaces the degraded thiazole moiety attached to the HMP with a hydroxyl group, completely restoring HMP (15). This can then be salvaged and used in thiamin synthesis. Thiaminase I (ThiA) enzymes also catalyze the base substitution of thiamin with a nucleophile. Unlike TenA, ThiA is secreted (16) and utilizes a variety of organic compounds as nucleophiles for the base substitution (17-19), and our recent findings suggest that it too plays a role in thiamin salvage, as the HMP and thiazole from thiamin degradation can be salvaged. Further, it appears that it can salvage the HMP from the degradation of the thiamin analog, pyriethiamin, suggesting it may function on natural thiamin analogs and degradation products encountered in nature.

Another distinguishing characteristic is that *thiA* is not widely distributed as it is only found in a small subset of bacteria (1) and 2 amoeba (20). Of these microorganisms, a few of them are known human pathogens including *Clostridium botulinum*, *Burkholderia pseudomallei*, and *Naegleria fowleri*, but the enzyme is also found in their non-pathogenic close relatives *Clostridium sporogenes*, *Burkholderia thailandensis*, and *Naegleria gruberi*. *Paenibacillus thiaminolyticus*, *Paenibacillus apiarius*, and *Paenibacillus dendritiformis* also produce this enzyme. Thiaminase I production by *C. sporogenes* has been implicated in thiamin deficiency in ruminants (21), and *P. thiaminolyticus* has been linked as one of the potential sources of

thiaminase I that causes the thiamin deficiency malady, early mortality syndrome, in salmonid fishes (22). However, *P. thiaminolyticus* is not always isolated from fish with high thiaminase I activity (23), making the link somewhat tenuous. Bacterial production of thiaminase I may also result in human thiamin deficiency; *P. thiaminolyticus* was originally isolated from the feces of humans experiencing thiamin deficiency (17, 24), and recently a case of suspected infant botulism was cured through thiamin supplementation. Thiamin deficiency and botulinum toxin both induce paralysis, suggesting that the paralysis occurring in the infant was exacerbated due to ThiA produced by *C. botulinum* (25). Although not pathogenic to humans, *B. thailandensis* is able to infect mammalian models at high doses (26, 27) and insect models (28, 29). Both *P. thiaminolyticus* and *P. apiarius* have been isolated from dead honeybees (30, 31), though they were not believed to be the causative agent of death. All these factors, taken together with the fact that animals cannot synthesize their own thiamin, and HMP has been shown to be toxic to animals (32), suggests a potential role for the enzyme in pathogenesis or thiamin competition with animal hosts.

The enzyme may also play a role in competition for thiamin amongst microorganisms. The gene is located in a conserved operon that encodes for the production of the HMP antimetabolite bacimethrin (33). When bacimethrin is combined with thiazole, it forms 2-methoxythiamin, which upon binding to proteins that require TPP as a cofactor, renders them non-functional, effectively killing cells (34). Due to its secreted nature, ThiA likely degrades environmental thiamin, and ThiA producers are able to recycle and grow on these constituents. This may provide the producer with a competitive advantage against other microorganisms as this reduces the amount of environmental thiamin available for competitor consumption, while still putting the compound in form it can use. This would likely be extra potent against thiamin auxotrophs

that rely on the import of environmental thiamin for their survival, as the already scarce essential nutrient would become even scarcer. Further, by depleting environmental thiamin, it might make bacimethrin more potent as competitors salvaging HMP due to the lack of environmental thiamin, may be more likely to take up this antimetabolite, incorporate it and die. We therefore hypothesized that ThiA may not only function as a competition factor against animals, but also against bacteria.

To address the first hypothesis, we used a *Drosophila melanogaster* infection model and multiple ThiA producing bacteria. We performed septic injury infections on adults, as well as oral infections on adults and larvae. A subset of infections were performed with *B. thailandensis* as we have both a wild-type (E264) and thiaminase I⁻ (BT10432) strain, as well as thiamin auxotrophic mutants generated in both backgrounds. Another subset of infections were performed with *P. apiarius*, *P. dendritiformis*, and *P. thiaminolyticus*, as well as a group of their non-ThiA producing close relatives, and other Paenibacilli. In the *B. thailandensis* infections, we found no ThiA effect, as both non-auxotrophic strains were highly infectious. In the thiamin auxotrophic strains there appeared to be an advantage with having a thiaminase I as the BT10432-*thiC* survival curve is more delayed than the E264-*thiC* infections. This was not significantly different. In the Paenibacilli infection subset, the ThiA producing bacteria were all highly pathogenic, in contrast to their Paenibacilli cousins, which demonstrated different degrees of pathogenicity. This suggests there is a correlation between having a ThiA and being pathogenic, but due to the lack of thiaminase I⁻ mutants in the Paenibacilli, we do not have the ability to determine if it is causal. To test the second hypothesis, we utilized *thiA*⁺ and *thiA*⁻ *B. thailandensis* strains and performed both in culture and supernatant competition experiments with artificial and natural thiamin auxotrophs. Our competition results did not clearly indicate

evidence for ThiA serving as a competition factor in the conditions we tested, but it does seem to sensitize competitors to bacimethrin.

A.2 MATERIALS AND METHODS

***Drosophila melanogaster* septic injury infections.** Canton^S flies were used for septic injury infections, and fly stocks were maintained on a standard diet containing 5% (w/v) yeast, 4% (w/v) glucose, 6% (w/v) cornmeal diet at 25°C (35). For *B. thailandensis* strain infections, *B. thailandensis* E264, BT10432, E264-*thiC* (E264 Δ *thiC*::*tetR*) and BT10432-*thiC* (BT10432 Δ *thiC*::*tetR*) were used. *B. thailandensis* cells were grown overnight in Tryptic Soy Broth (TSB) at 27°C with shaking at 200 rpm. For *Paenibacillus* species infections almost all species were grown overnight, except where noted. *Paenibacillus alvei* USDA B-383 and *Paenibacillus larvae* subsp. *pulvifaciens* USDA B-3685 were grown at 30°C in Nutrient Broth with shaking at 225 rpm. *P. apiarius* NRRL B-23460 was grown in Tryptone Glucose Yeast extract broth (TGY) at 30°C with shaking at 225 rpm. *Paenibacillus popilliae* NRRL B-2309 and *Paenibacillus lentimorbus* NRRL B-2522 were grown 2-3 days in advance of infection in MYPGP broth at 30°C with shaking at 200 rpm. *Paenibacillus dendritiformis*, and *Paenibacillus thiaminolyticus* NRRL B-4156 were grown in TSB at 37°C, while *Paenibacillus azotofixans* ATCC 35681 was grown in TSB at 30°C, all with shaking at 225rpm. *Paenibacillus macerans* BICM B-51 and *Paenibacillus polymyxa* ATCC 842 were grown in Lysogeny Broth at 37°C with shaking at 225 rpm. Cells were pelleted and washed twice with 1x PBS buffer and resuspended in 1x PBS to a final concentration of OD₆₀₀ = 1. For infections, flies were anaesthetized on CO₂ pads and infected through the pricking method (36), in which a sterile, sharpened needle was dipped into the bacterial suspension and punctures the thorax to deliver the bacteria. Mock-infected controls

were infected with sterile 1x PBS. After 2 hours, flies were checked for death. If death was observed, these flies were discarded from the growth curve as their death was a result of improper infection technique. Flies were checked periodically for death. Every 3 days, live flies were transferred to new food tubes. For all infections ~22 males were used per tube. For *B. thailandensis* infections, 3 tubes were used and this was replicated 3 times for a total of ~180 fly infections for each strain. For *Paenibacillus* species infections, 2 tubes were used and this was replicated 5 times for a total of ~200 fly infections for each strain. 3-5 day old adult *spätzle* male flies (*spz^{rm7}/TM6C*) were also infected with *Paenibacillus* species (excluding *P. popilliae* and *P. lentimorbus*) the same way as described. 2 tubes were used per replicate with 2 replicates occurring. These flies were maintained the same way as the Canton^S.

***B. thailandensis* infection statistics.** The experiments ran until every infected fly died, or for 216 hours. The data for each experiment was pooled and analyzed via a Cox Proportional Hazards (CoxPH) model in SAS 9.4. The thiamin capable strains were compared to each other, while the thiamin auxotroph strains were compared to each other. Infection type and replicate were fixed effects, with the tube in each experiment being a random effect.

Madagascar hissing cockroach infections. Madagascar hissing cockroaches were infected with both the E264 and BT10432 *B. thailandensis* strains, as well as all 3 *thiA* containing *Paenibacillus* species. We followed the protocol generated by Fisher and colleagues, in which 1.5 to 2 inch juveniles were inoculated with 25µL of OD₆₀₀ = 0.1 of bacterial suspension (prepared the same way as with *D. melanogaster* infections), through injection between the third and fifth terga of the abdomen (28).

Thiaminase activity. Flies that had just died were taken and frozen immediately at -80°C. Dead flies infected with the same species were pooled and weighed. 100 mM Phosphate Buffer pH 6.5 was added at 2.5x concentration (w/v). Samples were homogenized, and spun 2x to clarify the solution. This was then used in the standard thiaminase I assay developed by Kraft et al. (37). Madagascar hissing cockroach hemolymph was also assessed for thiaminase I activity after death from infection. Cockroaches were immediately frozen, and thawed for processing. Cockroach abdomens were stabbed with a sterile needle, and squeezed into a sterile microfuge tube to collect the hemolymph. The thiaminase I assay was conducted using this hemolymph.

Adult oral infections. Oral infections were conducted using the E264 and BT10432 strains of *B. thailandensis* as well as *P. alvei*, *P. apiarius*, *P. azotofixans*, *P. dendritiformis*, *P. larvae* subsp *pulvifaciens*, *P. macerans*, *P. polymyxa*, and *P. thiaminolyticus*. Strains were grown the same way as in the previous section, except large volumes were used for growth (between 50-100mLs), as the cells were pelleted, washed, and concentrated to $OD_{600} = 200$ in 1x PBS. 3-5 day old adult female Canton^S flies were anaesthetized with CO₂, then starved for 2 hours at 29°C. After the starvation period ended, the concentrated bacterial suspensions were mixed at a 1:1 ratio with 5% (w/v) sucrose, giving the final $OD_{600} = 100$ (38). 150μL of this solution was added to sterile filter paper placed directly on the surface of fresh standard fly food containing glucose. The starved females were added to this food, and death was observed after 2 hours. Those that died were removed from the study. Death was observed every couple hours for a week.

Larval oral infections. Adult male and female *D. melanogaster* Canton^S flies were placed in laying cages containing grape agar and laid eggs overnight. Eggs were then collected and ~30 eggs were placed on standard diet containing 5% (w/v) yeast, 4% (w/v) sucrose, 6% (w/v) cornmeal, and 2-3 drops of blue food coloring at 25°C. After 40 hours post egg transfer, eggs were observed to see if they hatched. Unhatched eggs were recorded and subtracted from the study. Bacterial cells were grown the same way as described in the septic injury infections section. This time, they were resuspended in 1x PBS to an OD₆₀₀ of 30. 150µL of this cell suspension was added to the food surface. Pupal formation and adult emergence were recorded. N = ~90 eggs for each bacterial treatment.

Screening of antibiotic sensitivity for *thiA* containing *Paenibacilli*. *P. apiarius*, *P. dendritiformis*, and *P. thiaminolyticus* were grown in TSB at their respective temperatures overnight, and subcultured at an OD₆₀₀ = 0.025 in TSB containing either ampicillin, carbenicillin, chloramphenicol, gentamicin, spectinomycin, or trimethoprim at 100, 50, 25, 10, or 5µg/mL.

Generation of gene disruptions. Based on antibiotic sensitivities of *thiA* containing *Paenibacillus* species, plasmids were constructed to partially delete and disrupt the *thiA* gene in each species. Plasmids were also constructed to partially delete the *bcmA* gene in *B. thailandensis* E264 following the methods detailed in (39, 40). This protocol utilizes a uracil auxotroph *Saccharomyces cerevisiae* mutant strain to recombine PCR fragments into the linearized pMQ87 vector, which contains the URA3 gene to complement the auxotrophy, and a gentamicin resistance gene. This plasmid can replicate in both *S. cerevisiae* and *E. coli*. In each

step, the PCR products were purified using the QIAquick PCR Purification Kit (QIAGEN). Briefly, the 5' end of the *thiA* genes were amplified from each *Paenibacillus* species genomic DNA, using the pMQ87-*thiA* 5' species specific primer set with the forward primer containing a 35bp overlap with pMQ87 plasmid. The 3' end of *thiA* was amplified using the pMQ87-*thiA* 3' primer set, with the reverse primer containing 35bp overlapping with pMQ87. The *thiA* genes were disrupted with a tetracycline resistance gene amplified from *B. thailandensis* BT02155 genomic DNA, which contains the T8 transposon with a *tetR* gene (41) for all 3 species, or with the ampicillin resistance gene from pUC19. The *P. apiarius thiA* was also disrupted with the spectinomycin resistance gene from pOM1(42). The *P. dendritiformis thiA* was also disrupted with the trimethoprim resistance gene from *B. thailandensis* BT10432 genomic DNA. The *P. thiaminolyticus thiA* was also disrupted with the *cmR* (*cat*) amplified from pBeloBAC11. The *bcmA* gene from *B. thailandensis* E264 was disrupted with the *kanR* from pUC19. To amplify the antibiotic resistance cassettes, the forward primer contained a 42bp overlap with the 3' end of the 5' fragment of the gene to be disrupted, and the reverse primer contained a 42bp overlap with the 5' end of the 3' fragment of the gene to be disrupted. These fragments were combined with the uracil auxotroph yeast, and linearized pMQ87, and grown on media lacking uracil. Plasmids were extracted from transformant colonies and DH5 α *E. coli* was transformed with the plasmids following the chemical transformation protocol of (43), and plated on 10 μ g/mL gentamicin containing LB plates. Transformants were screened using colony PCR to ensure that the construct was the expected size, and if so the plasmids were extracted.

***Paenibacillus* thiaminase I containing species transformation attempts.** Three transformation protocols were used, the first was generated for the transformation of *Paenibacillus alvei* CCM

2051T, detailed by Zarschler et al. (44). This protocol generated electrocompetent by washing cells 5x in ice cold electroporation buffer (250 mM sucrose, 1 mM HEPES, 1 mM MgCl₂, and 10% glycerol, pH 7.0) and resuspending them in the buffer at 1/500x initial culture volume. Electrocompetent cells were mixed with either purified PCR products containing the disruption constructs, circular vector, or linearized vector containing the disruption constructs. The electroporation was carried out as described in (44), using a BioRad Gene Pulser electroporator. Immediately after electroporation, cells were mixed with pre-warmed TSB and incubated for either 2, 6, 18, or 24 hrs and plated on TSB containing the necessary antibiotic for the marker. The next protocol attempted was generated for *P. larvae* (45). The electroporation buffer used in this protocol consisted of 0.625 M sucrose and 1 mM MgCl₂. Upon electroporation, cells were directly incubated in pre-warmed TSB and grown for 2, 6, 18, or 24hrs and then plated. We also tried the protocol generated for *P. polymyxa* (46). This protocol calls for denaturing DNA, as this may improve transformation efficiency, to do this, DNA was denatured in 0.2 mM EDTA and 0.2 M NaOH, or through UV treatment (46). The transformations were plated on TSA plates containing the specific antibiotic at various concentrations, ranging from 10, 25, 50, 100, and 200 µg/mL. Potential transformants were screened via colony PCR using confirmation primer sets.

Generation of *bcmA* mutants in *B. thailandensis*. To generate the *B. thailandensis* disruption mutants, we followed the protocol generated by Thongdee and colleagues, which takes advantage of *B. thailandensis*' natural competence (47). Competent E264 and BD20 ($\Delta btaK$) (48) cells were generated as described (47). The BD20 cells were kindly given to us by Dr. Josie Chandler. Competent E264-*thiC* cells were generated as described (47). The *bcmA* disruption

fragment was PCR amplified from pMQ87-*bcmA-kanR* and added to the competent E264 strain to generate E264-*thiCbcmA*. After 48 hrs of incubation, the E264 transformation culture was plated on LB with 100 µg/mL kanamycin. To generate BD20-*thiA* and BT10432-*thiCbcmA*, genomic DNA extracted from BT10432 was added to BD20 and E264-*thiCbcmA* cells respectively. The BD20 transformation cultures were plated on LB with 100 µg/mL trimethoprim. The E264-*thiCbcmA* transformation was plated on LB with 50 µg/mL of tetracycline and 100 µg/mL kanamycin. The BT10432-*thiCbcmA* cells were plated on LB with 50 µg/mL of tetracycline, 100 µg/mL trimethoprim and 100 µg/mL kanamycin. Transformants were screened via colony PCR to ensure the constructs recombined as well as to ensure no homologous recombination occurred at the other mutated loci.

Supernatant and pure thiaminase I competitions. E264-*thiCthiG* and BT10432-*thiCthiG* were grown in TSB at 27°C with shaking at 200 rpm. Sample was collected at 12, 24, 36, 48, and 60hrs and centrifuged. The supernatant was sterilized through filtration using a 0.2µm filter, and stored at -80°C. The sterile filtrate from each time point was mixed with sterile TSB at a final volume of 10% (v/v). This was then incubated for 3hrs at 27°C with shaking at 200 rpm. The natural thiamin auxotrophs *Aerococcus viridans* USDA B-2309, *Lactobacillus plantarum* DmCS_001 (49), *P. alvei*, and *P. macerans* were individually added to this media at OD₆₀₀ = 0.025. *L. plantarum* and both *Paenibacillus* species were grown in TSB, with *L. plantarum* growing at 30°C with no shaking, and the two *Paenibacillus* growing at 37°C with shaking at 225 rpm. *A. viridans* was grown at 27°C with shaking at 200 rpm. Growth was observed at 18 and 24 hrs post inoculation. They were also treated with sterile supernatant from BD20 and BD20-*thiA* TSB cultures, collected at 12, 24, 36, and 48 hrs in the same manner. These

supernatant experiments were performed in the same manner as the others. For pure thiaminase I treatments, purified BcmE (50) was added at a concentration of 50µg/mL to TSB. The TSB was incubated at 37°C for 3 hours, and then natural thiamin auxotrophs were added to BcmE treated media at an OD₆₀₀ = 0.025, and incubated as described earlier.

Head to head competition experiments. BD20 were competed against *E. coli thiC* mutants (*thiC* 765Δ::kan), purchased from the *E. coli* Genetic Stock Center at Yale University in a head to head competition. The experiment was conducted as outlined by Chandler et al. (48). BD20, BD20-*thiA*, and *E. coli thiC*⁻ pure cultures were grown in TSB overnight, then subcultured in fresh TSB at an OD₆₀₀ = 0.05 and grown for 3 hrs. BD20 strains were combined in 10mL of TSB with the *E. coli thiC*⁻ mutant so that there was an initial 10:1 ratio of *B. thailandensis* to *E. coli* (starting OD₆₀₀ of 0.05:0.005) in 3 replicates for each competition. The cultures were grown at 27°C with shaking at 200 rpm. Samples were plated right after inoculation (0 hr), as well as at 24 and 48 hrs on TSA at 37°C with a WASP2 Spiral Plater (Microbiology International). Colonies were distinguished based on morphology and CFUs determined based on the spiral plater manual. The *B. thailandensis* competitions with *P. macerans* were carried out the same way. E264-*thiC*, E264-*thiCbcmA*, BT10432-*thiC*, and BT10432-*thiCbcmA* were competed against *P. macerans* in both a 1:1 and 5:1 *B. thailandensis* to *P. macerans* ratio. The 1:1 ratio competitions were plated at 0, 12, and 24 hrs, while the 5:1 ratio competitions were plated at 0, 6, and 12 hrs. Competitions were assessed for statistical significance using JMP 12.0.1, via a fit model including a Least Squares Means test. Replicate was used as a random effect, with competition, time, and the cross of competition*time as fixed effects. Competitions were Bonferroni corrected for the comparisons tested.

Table 1 Primers used in this study

Name	Primer
pMQ87-bcmA 5' F	AGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCGAAGACAGCATCAACCTG
bcmA 5' F	GAAGACAGCATCAACCTG
bcmA 5' R	AAGATCTGCAGCACGACG
bcmA 3' F	CTCTTCTGGATCTACCGC
bcmA 3' R	AAGCCGAGTTCCATCGTC
pMQ87-bcmA 3' R	AAACAGCTATGACCATGATTACGAATTCGAGCTCGAAGCCGAGTTCCATCGTC
bcmA-kanR F	CCGTACTGGCGCGACGTGCTCGGCCTCTTCTGGATCTACCGCAGGATCTGATGGCGCA GG
kanR-bcmA R	GGCGGGTTGCCCCCCTTGATTTCGCGGTAGATCCAGAAGAGCATCAGGAAATTGTAA GCG
bcmA-kanR construct confirmation F	CGACGACGCGACGGAAGT
bcmA-kanR construct confirmation R	TTCGCCGCCTTGAACAGC
Apiarius pMQ87-thiA 5' F	AGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCTCGGGTCATGTCGAGCTG
Apiarius thiA 5' F	TCGGGTCATGTCGAGCTG
Apiarius thiA 5' R	GACGTCAGGCTCCTTCCG
Apiarius thiA 3' F	AGACATCCCATCCAGAC
Apiarius thiA 3' R	TACCAGATCGCAGCTGTC
Apiarius pMQ87-thiA 3' R	AAACAGCTATGACCATGATTACGAATTCGAGCTCGTACCAGATCGCAGCTGTC
Apiarius thiA-tetR F	TTTGAAGCGGCCGTTCTCGAACAGTGGGAGCGGAAGGAGCCTGCTAACGCAGTCAGG CAC
Apiarius tetR-thiA R	CGACATTCGCTAGCTCTTTGGCTAGTCTGGGATGGGATGTCTTCCGTTAGCGAGGTGC CG
Apiarius thiA-ampR F	TTTGAAGCGGCCGTTCTCGAACAGTGGGAGCGGAAGGAGCCTGGGAAATGTGCGCGG AAC
Apiarius ampR-thiA R	CGACATTCGCTAGCTCTTTGGCTAGTCTGGGATGGGATGTCTGTAACTTGGTCTGAC AG
Apiarius thiA-specR F	TTTGAAGCGGCCGTTCTCGAACAGTGGGAGCGGAAGGAGCCTAGTGGCGGTTTTCAT GGC
Apiarius specR-thiA R	CGACATTCGCTAGCTCTTTGGCTAGTCTGGGATGGGATGTCTGCTTGAACGAATTGTT AG
Apiarius construct confirmation F	GCCGTGACGCGAATTGTG
Apiarius construct confirmation R	GACTCTATTGCCTCTTCG

Dendritiformis pMQ87-thiA 5' F	AGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCGAGAGCGTTGAGGAGGCG
Dendritiformis thiA 5' F	GAGAGCGTTGAGGAGGCG
Dendritiformis thiA 5' R	ACTCCAGCTTGACGCCAG
Dendritiformis thiA 3' F	CACAGTATCTGTTGCCTGC
Dendritiformis thiA 3' R	CGCAACGCTTCATTCCAC
Dendritiformis pMQ87-thiA 3' R	AAACAGCTATGACCATGATTACGAATTCGAGCTCGCGCAACGCTTCATTCCAC
Dendritiformis thiA-tetR F	TCCTTGACCAGTGGCAGCGCCAGGAGCCTGGCGTCAAGCTGGGCTAACGCAGTCAGG CAC
Dendritiformis tetR-thiA R	ATCAATGCCTCATATACCTGATGCCGGGCAGGCAACAGATACTCCGTTAGCGAGGTG CCG
Dendritiformis thiA-ampR F	TCCTTGACCAGTGGCAGCGCCAGGAGCCTGGCGTCAAGCTGGGGGAAATGTGCGCGG AAC
Dendritiformis ampR-thiA R	ATCAATGCCTCATATACCTGATGCCGGGCAGGCAACAGATACGTAACTTGGTCTGA CAG
Dendritiformis thiA-tmR F	TCCTTGACCAGTGGCAGCGCCAGGAGCCTGGCGTCAAGCTGGGGCATCCAAGCAGCA AGC
Dendritiformis tmR-thiA R	ATCAATGCCTCATATACCTGATGCCGGGCAGGCAACAGATACAACCGGGCAGGCCAT GTC
Dendritiformis construct confirmation F	GGATGAGCATCTGGAGCG
Dendritiformis construct confirmation R	GAAGATGTTTCAGCGCCAC
Thiaminolyticus pMQ87-thiA 5' F	AGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCAAGCGGTTGGTCGGCCTG
Thiaminolyticus thiA 5' F	AAGCGGTTGGTCGGCCTG
Thiaminolyticus thiA 5' R	ACGCCAGGCTCTAGCTGC
Thiaminolyticus thiA 3' F	GGATACGGTAGAGCAAGC
Thiaminolyticus thiA 3' R	AAATGCGACGGTGCCGG
Thiaminolyticus pMQ87-thiA 3' R	AAACAGCTATGACCATGATTACGAATTCGAGCTCGAAATGCGACGGTGCCGG
Thiaminolyticus thiA-tetR F	CCAAGCAGCCGTCCTTGACCAGTGGCAGCAGCTAGAGCCTGGGCTAACGCAGTCAGG CAC
Thiaminolyticus tetR-thiA R	GATATTGGCCATCGGCCTGCGGGCGCAGAGCTTGCTCTACCGTCCGTTAGCGAGGTG CCG
Thiaminolyticus thiA-ampR F	CCAAGCAGCCGTCCTTGACCAGTGGCAGCAGCTAGAGCCTGGGGGAAATGTGCGCGG AAC
Thiaminolyticus ampR-thiA R	GATATTGGCCATCGGCCTGCGGGCGCAGAGCTTGCTCTACCGGTAACTTGGTCTGAC AG

Thiaminolyticus thiA-cat F	CCAAGCAGCCGTCCTTGACCAGTGGCAGCAGCTAGAGCCTGGTCTTCAACTAAAGCA CCC
Thiaminolyticus cat-thiA R	GATATTGGCCATCGGCCTGCGGGCGCAGAGCTTGCTCTACCGGTAGAGGATCTGGAG CTG
Thiaminolyticus construct confirmation F	GCAAGACGCCCTATATCG
Thiaminolyticus construct confirmation R	AGGAGATGTGCCGACCAC

A.3 RESULTS

ThiA does not provide a significant advantage to *B. thailandensis* during *D. melanogaster* infection

Pilátová and Dionne recently demonstrated that *B. thailandensis* E264 is pathogenic to wild-type *D. melanogaster* Oregon^R. They displayed that this killing and proliferation inside the host was independent of the presence of their type III and their type VI secretion systems, as mutants without these systems behaved like E264 infections (29). Further, the team demonstrated that *B. thailandensis* conditioned media was lethal to flies when injected into them. These led us to hypothesize that ThiA may be a factor contributing to this lethality as it is secreted through the general secretory pathway. To address this, we infected *D. melanogaster* through septic injury with BT10432 and E264 *B. thailandensis* strains (Figure 1). The survival curves showed that both strains kill *D. melanogaster* effectively, generally within 48 hours post infection, and there is no statistical difference between the survival curves of the infections with the 2 strains (Figure 1). E264-infected flies collected after their death showed consistent ThiA activity when screened, suggesting that the enzyme is active during infection, however, it is not significantly contributing to the pathogenicity. There was also no difference in pathogenicity between the 2 strains in infections of immunodeficient *relish* flies, or by lowering the starting OD₆₀₀ to 0.1 (data not

shown). Our previous studies depicted that ThiA gives *B. thailandensis* thiamin auxotrophs the ability to grow in media when *thiA*⁻ thiamin auxotrophs are unable. We performed infections using the E264-*thiC* and BT10432-*thiC* strains to see if this growth advantage would occur during infection inside *D. melanogaster*. The auxotrophs killed at a much slower rate than their thiamin capable counterparts, as the killing is generally delayed by 5 days. The survival curve shows that the E264-*thiC* cells appeared to kill a little faster than the BT10432-*thiC* cells (Figure 1 dashed lines), however, the survival curves are not significantly different from one another ($p = 0.0561$). This suggests that in these auxotrophs, ThiA may be playing a small role in the pathogenicity and acquisition of thiamin, however, the role is not very significant during *D. melanogaster* infection.

There is a correlation between ThiA production and pathogenicity in Paenibacilli

We next set out to determine if the ThiA producing *Paenibacillus* species were also pathogenic to *D. melanogaster*. We performed septic injury infections with *P. apiarius*, *P. dendritiformis*, and *P. thiaminolyticus*, as well as a few known *Paenibacillus* insect pathogens. These included *P. alvei* which is a close relative of *P. apiarius* (51) and infects and produces signs of European foulbrood in honeybees (52), *P. larvae pulvifaciens* which is responsible for powdery scale in honeybees (53), and *P. popilliae* and *P. lentimorbus* which are responsible for Milky Spore disease in Japanese beetles (54). The latter 2 species are close relatives to *P. thiaminolyticus* and *P. dendritiformis*. Infections were also performed using soil and rhizosphere *Paenibacillus* species including *P. azotofixans* (55), *P. macerans* (56), and *P. polymyxa* (57) to determine the range of infectivity of Paenibacilli. The survival curves depicted in Figure 2A demonstrate that the ThiA producing bacteria kill the most amount of flies, and there were

differences in the kinetics of killing. *P. apiarius* was the most efficient killer, typically killing flies within 48 to 52 hrs. *P. thiaminolyticus* also killed all the flies it infects, but at a slightly slower rate, usually within 72 hrs. The *P. dendritiformis* infection has a slower kinetic, killing most of the flies within 120 hrs. Unlike the other *thiA* Paenibacilli, *P. dendritiformis* did not kill 100% of the flies it infected, rather it killed about 95%. Further, all 3 species produced ThiA during infection inside the fly. Of the non thiaminase I producers, *P. alvei* was the most effective killer, as ~50% of the flies it infected die. *P. polymyxa* and *P. macerans* were the only other infections that showed intermediary killing, whereas the other species kill no more than the mock-infection control. These results demonstrated that the *thiA* containing Paenibacilli are highly pathogenic to *D. melanogaster* during septic injury infection and vary in infection kinetics. In contrast, non thiaminase I producers are not as infective. Unfortunately, all attempts to transform and knock out thiaminase I production in the 3 species failed, so we do not have a good control to determine if ThiA is the main culprit in their infectivity, but results suggest its production may be correlative.

In *D. melanogaster*, the Toll pathway is the signaling cascade that elicits the immune response to combat Gram-Positive bacterial infections (58). The cytokine Spätzle stimulates the activation of the Toll immune response, and *spätzle* mutants are more susceptible to infection by Gram-Positive bacteria (59). We wanted to determine if the killing kinetics would change when immune-deficient flies were infected with the Paenibacilli species. We found that when *spätzle* flies were infected with the Paenibacilli that produce an intermediary infective phenotype, they were more susceptible to death than wild-type Canton^S flies (Figure 2B). *P. alvei* was able to kill 95% of the flies it infected and *P. macerans* was able to kill 90%. The other species were still not infective, suggesting that *D. melanogaster* is not a suitable host for them. The same pattern was

not observed when infected with the *thiA* strains. The killing kinetics were not faster and appear to be more delayed than the infections in Canton^S (Figure 2C). The lack of an increase in killing speed suggests that the Toll immune response is not effective in controlling these invaders. This could be because their genomes appear to encode for meso-diaminopimelate containing peptidoglycan.

We also found that both BT10432 and E264 were pathogenic to Madagascar hissing cockroaches in a similar fashion (Figure 3). This suggested that for *B. thailandensis* ThiA was not playing a significant role. The *thiA* *Paenibacillus* species are pathogenic to different degrees. *P. apiarius* was the most pathogenic of all bacteria tested, killing 100% within 48 hrs. The *B. thailandensis* species killed 100% within 96 hrs. *P. thiaminolyticus* killed more than 80% of the cockroaches it infected by 120 hrs. *P. dendritiformis* killed over 60% of the cockroaches it infected. All *thiA* strains produced ThiA during infection within the host.

The *thiA* containing bacteria are not orally infective to *D. melanogaster* adults, and ThiA does not play a significant role in larval infections.

Our results from septic injury did not signify a clear link between ThiA and pathogenicity. We hypothesized that there may be a link through an oral infection route because if thiaminase I is degrading thiamin in the food and gut, it could possibly lead to thiamin deficiency and death. When flies were orally challenged with the bacteria used in the septic injury infections, no pathogenicity was observed (data not shown) indicating that these organisms are not pathogenic through this route. The lack of pathogenicity of *B. thailandensis* E264 is in contrast to what was observed by Pilátová and Dionne, as it was highly pathogenic in an oral route.

Our findings from another study demonstrated that *D. melanogaster* adults do not require thiamin in their diet, and that may partially account for the lack of pathogenicity in the adult infections. Larval development is energetically demanding in *Drosophila* (60, 61) and is heavily dependent upon the availability of dietary nutrients (62-64), therefore we surmised that the larval stage may give us a thiaminase I effect through oral infection. To test this, we added a concentrated solution of the bacterial species tested in the previous sections (excluding *P. popilliae* and *P. lentimorbus*) to larvae feeding in the L2 stage of development, and observed their ability to pupate and form adults. We found that only when challenged with *B. thailandensis*, significant death occurred (Figure 4). This death was irrespective of if *thiA* was present in the genome, as both E264 and BT10432 prevented pupation and adult eclosion from occurring in 100% of the larvae. None of the Paenibacilli killed larvae significantly different from the uninoculated control. This suggests that in these conditions, thiaminase I does not play a role in the larval infections.

Bactobolin A *B. thailandensis* mutants did work as competitors for investigating thiaminase I's role in bacterial competition

Due to thiaminase I's ability to degrade thiamin and for its producers to salvage the thiamin breakdown products we hypothesized that it would likely give its producers a competitive advantage against other bacteria, in particular thiamin auxotrophs. To test this, we performed different competition experiments with artificial and natural thiamin auxotrophs. For these experiments we used *B. thailandensis* as the ThiA competitor due to having the BT10432 *thiA*⁻ control. *B. thailandensis* produces the antibiotic bactobolin A, which can confound competition experiments, therefore we conducted our first set of competitions in the BD20

($\Delta btaK$) background, deficient in bactobolin A production (48). We also generated a *thiA*⁻ in this background as well. We competed both strains against an *E. coli* MG1655 *thiC*⁻ mutant in a head-to-head competition (Figure 5). The experiments showed that neither *B. thailandensis* strain was highly competitive against *E. coli*, as the ratio of *B. thailandensis* to *E. coli* continued to decrease as time went on despite starting a 10-fold higher density. There was no difference between having a thiaminase I and not having it. These results suggest that ThiA plays no role in the competition between *B. thailandensis* and thiamin auxotrophic *E. coli*. Further it demonstrated that *E. coli* likely was not the best candidate to utilize in competition experiments. We then decided to use 4 natural thiamin auxotrophs as competitors, after investigating their genomes for thiamin biosynthetic capabilities. *Aerococcus viridans* served as the first competitor as it lacks the genomic potential to create both moieties and does not have the *ykoCDEF* HMP transport system. *Lactobacillus plantarum* served as the second competitor, which lacks a *thiC*, *thiG*, and *ykoCDEF* system. *P. alvei* served as the third competitor which lacks both *thiC* and *thiG*, but has the *ykoCDEF* system. The fourth competitor was *P. macerans*, which can make the thiazole moiety, but lacks *thiC* and the *ykoCDEF* system. BD20 and BD20-*thiA* conditioned media were ineffective at limiting growth of the competitors, as all 4 competitors grew in the conditioned media despite the presence of ThiA (Table 2). These results suggest that in these conditions, thiaminase I is not enough to limit the growth of the competitors, and bactobolin production may be necessary to make *B. thailandensis* competitive.

Thiaminase I may sensitize cells to bacimethrin

Due to a lack of competition in the BD20 backgrounds, we shifted back to performing competitions in the *btaK*⁺ backgrounds. To prevent any thiamin or thiamin precursor production

from affecting results, we utilized *B. thailandensis* E264-*thiCthiG* and BT10432-*thiCthiG* mutants as competitors. We found that when *A. viridans* was grown in the *B. thailandensis* conditioned media, it could only grow in the media treated with the supernatant collected after 12 hours of E246-*thiCthiG* growth and BT10432-*thiCthiG* growth (Table 3). All others prevented growth. This suggested that although effective at preventing the growth of *A. viridans*, thiaminase I does not play a role. Supernatant collected at 12, 24, and 36 hrs from both *B. thailandensis* strains was ineffective at preventing the growth of *L. plantarum*, but the supernatants from the latter stages of the cultures were, in a ThiA independent fashion (Table 3). *P. alvei* was able to survive in all treatments. There appeared to be a ThiA effect on the inhibition of the growth of *P. macerans*. *P. macerans* treated with BT10432-*thiCthiG* supernatant was able to grow, however 18 hrs post inoculation into media containing the 12 hr E264-*thiCthiG* supernatant, it was unable to grow. Growth began to occur around 24 hours post inoculation. This demonstrated that ThiA treated media was transiently playing a role in the limiting of growth *P. macerans* at this stage. *P. macerans* was unable to grow in any of the other media treated with supernatant, suggesting that they were overcome with what was produced in the media by *B. thailandensis* later in its growth.

We hypothesized that ThiA sensitization to bacimethrin could be responsible for the transient nature of the growth prevention caused by the 12 hr E264-*thiCthiG* supernatant. This is because in the BT10432-*thiCthiG* treated TSB, ThiA was not present to degrade thiamin. This would have a two-fold effect; first there would be more thiamin available for *P. macerans* to salvage, and second there would not be as much HMP present in the media to potentially stimulate HMP uptake. When ThiA is present, there would be less thiamin available for uptake, but more HMP present. *P. macerans* would be more likely to import bacimethrin if it is trying to

salvage the HMP present in the media. The *P. macerans* cells may have been importing all the bacimethrin produced, and when the media was depleted of bacimethrin due to importation by *P. macerans*, those cells that did not import it were able to grow. *B. thailandensis* appears to have the genomic capabilities to make bacimethrin, though this has yet to be biochemically confirmed, and the biosynthetic genes are located in an operon containing *thiA*. To test this, we generated $\Delta bcmA::kanR$ mutants in the BT10432-*thiC* and E264-*thiC* backgrounds, and competed them against *P. macerans* in head to head competitions. The bacteria were first competed in a 1:1 ratio, in this experiment, at 24 hrs, the BT10432-*thiCbcmA* mutants had the lowest ratio of *B. thailandensis* to *P. macerans*, in comparison to BT10432-*thiC*, E264-*thiC*, and E264-*thiCbcmA* (Figure 6A). However, this difference was not significant. We repeated the competition, this time with a starting ratio of 5:1 *B. thailandensis* to *P. macerans* (Figure 6B). At 12 hours, we observed a significant difference in the ratio of *B. thailandensis* to *P. macerans* in the BT10432-*thiCbcmA* and E264-*thiC* competitions, as the BT10432-*thiCbcmA* was significantly lower ($p = 0.0013$). This finding supports the hypothesis that ThiA is sensitizing *P. macerans* to bacimethrin, as the cells are not as competitive when they lack both the ability to make bacimethrin and thiaminase I. The lack of statistical differences between any other comparison at the same time point shows that both *bcmA* and *thiA* are necessary to be most competitive.

Purified thiaminase I does not prevent growth of natural thiamin auxotrophs

To confirm if thiaminase I alone is a competition factor, we treated both TSB and defined media with purified thiaminase I and observed the growth of the 4 natural thiamin auxotrophs. All 4 species were able to grow in TSB pre-treated with and containing thiaminase I (Table 4), without any differences from growth in untreated TSB. It was also ineffective at preventing

growth when these already treated samples were serially transferred to more TSB pretreated with thiaminase I. This established that thiaminase I was not an effective competition factor in TSB. We then tried to utilize a completely chemically defined media to see if thiaminase I was effective at preventing growth, after serially transferring from thiaminase I treated TSB to the defined media. *A. viridans* was unable to grow at all in the media irrespective of thiaminase I treatment, while *P. alvei* could only grow in clumps in the media, irrespective of thiaminase I treatment. Both *L. plantarum* and *P. macerans* were able to grow at low densities in the media irrespective of thiaminase I being present (Table 4). These findings confirm that thiaminase I alone is not an effective competition factor against the thiamin auxotrophs in these conditions.

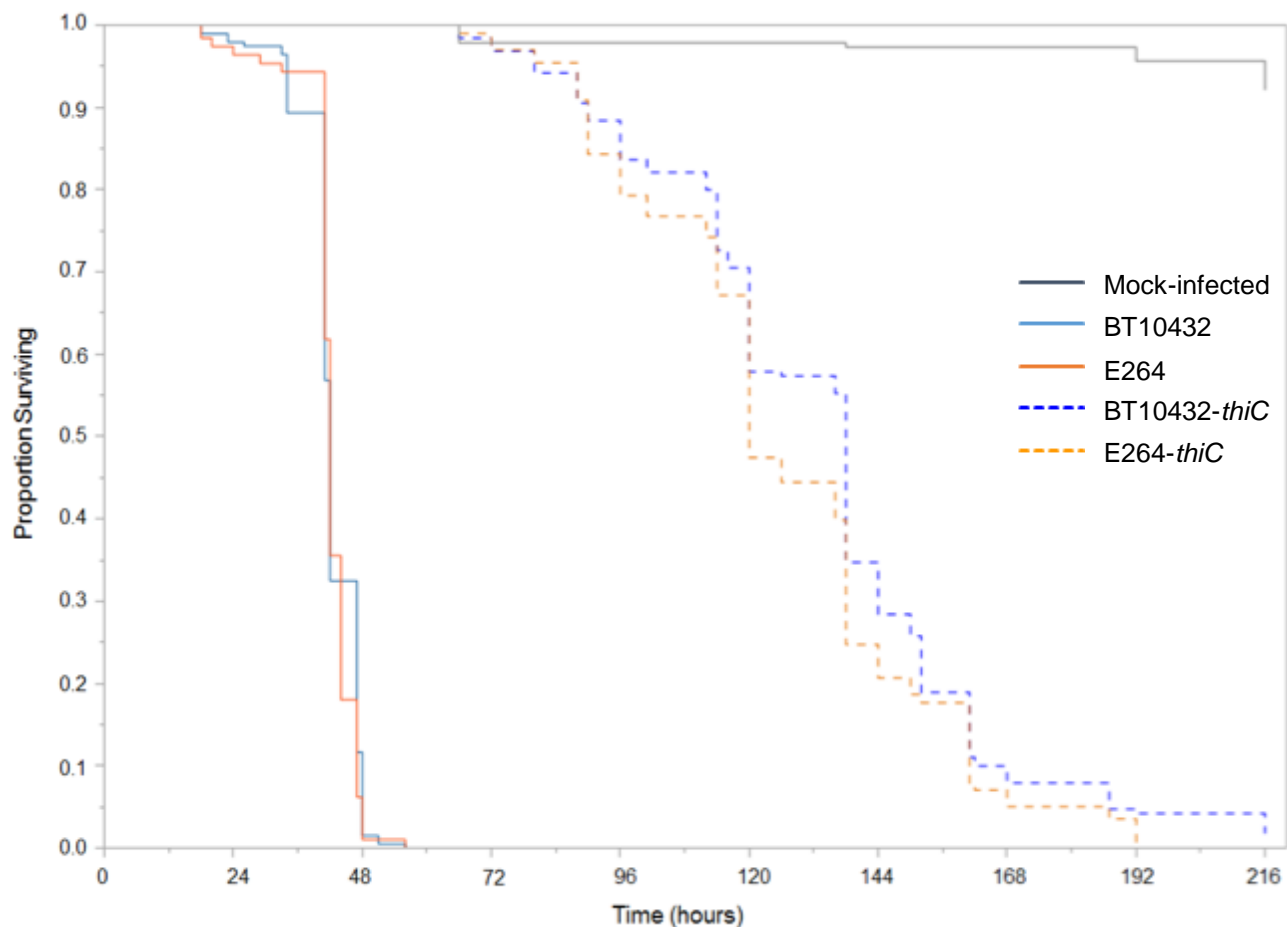


Figure 1 Survival curves of *B. thailandensis* infected *D. melanogaster*. Survival analysis shows that there is no statistical difference ($p = 0.3430$) between infections with the BT10432 and E264 strains as both kill almost all flies within 48 hrs of infection. The killing kinetics are delayed when infected with *thiC* mutants of both backgrounds, as the majority of killing is delayed by 5 days. The E264-*thiC* cells kill at a faster rate for most of the infection than the BT10432-*thiC* cells, but this is not statistically significant ($p = 0.0561$).

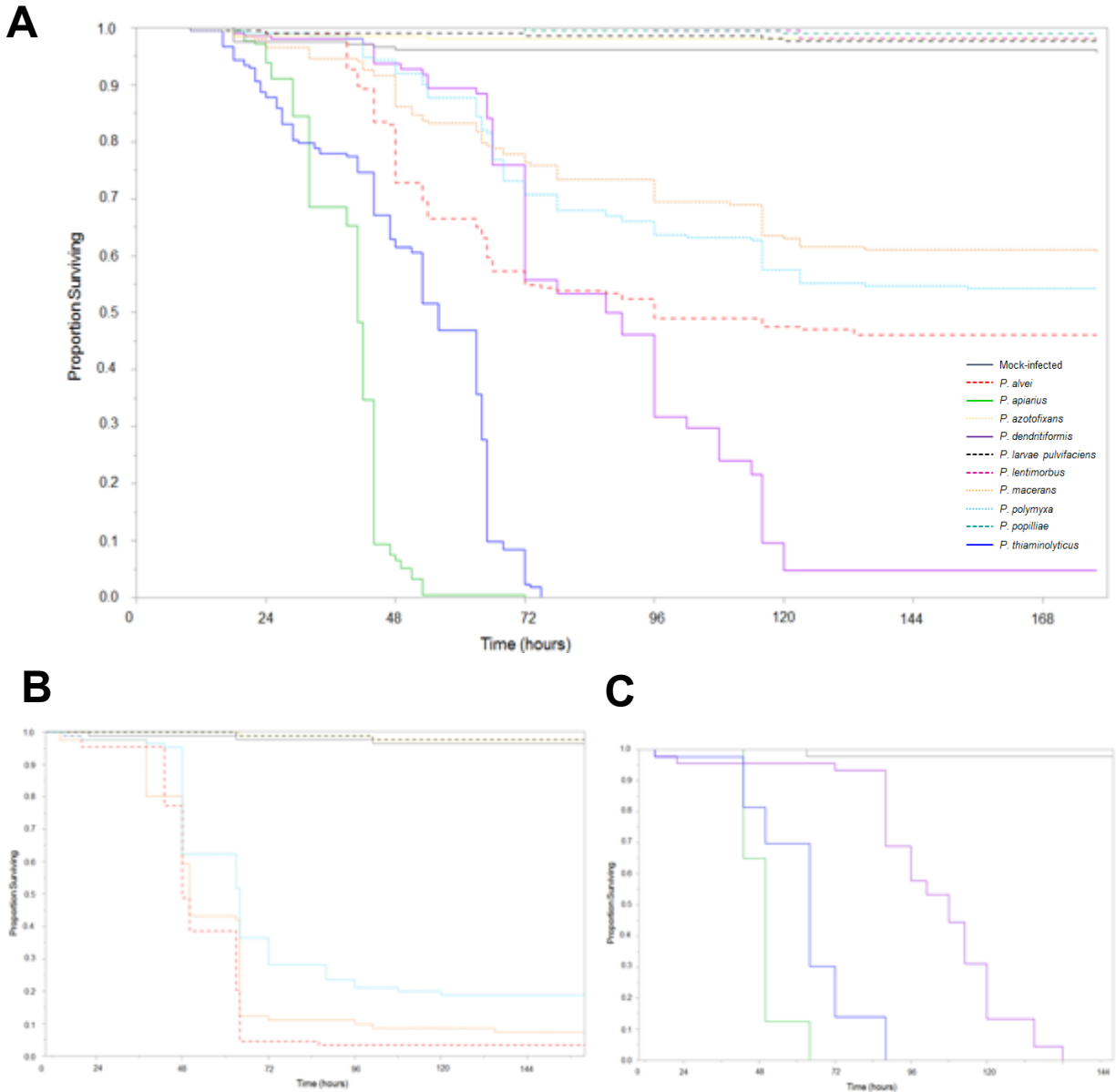


Figure 2 Survival curves of *Paenibacillus* species infected *D. melanogaster* strains. ThiA producers are represented in solid lines, known insect pathogens are represented by dashed lines, and rhizosphere/soil bacteria are represented by dotted lines. (A) is the survival curve of Canton^S flies infected with the *Paenibacillus* species. The three ThiA producers are the most infective with *P. apiiarius* killing almost all flies within 48 hrs and *P. thiaminolyticus* within 72 hrs. *P.*

dendritiformis kills at a slower rate, but kills 95% of all flies within 120 hrs. Of the non thiaminase I producers, *P. alvei* is the most effective pathogen, killing a little more than 50% of flies it infects, while *P. polymyxa* and *P. macerans* kill less than 50%. (B) depicts the non thiaminase I producers infecting *spätzle* flies. *P. azotofixans* and *P. larvae pulvifaciens* are still not pathogenic to these immunodeficient flies, however, *P. alvei*, *P. macerans*, and *P. polymyxa* are, as they all kill a higher proportion of flies then when they infect wild-type flies. (C) illustrates the infectivity of thiaminase I producing Paenibacilli in *spätzle* flies. With each species, the speed of killing is delayed in comparison to wild-type infections.

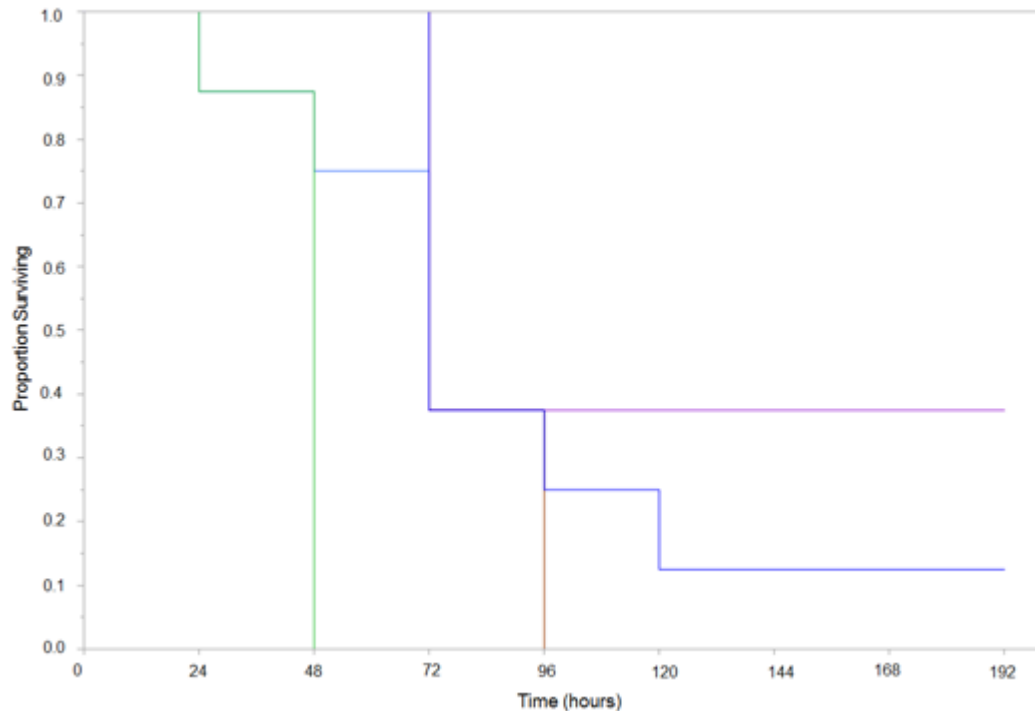


Figure 3 Infection of Madagascar hissing cockroaches with ThiA producing bacteria. ThiA does not play a significant role in the pathogenicity of *B. thailandensis* as there is no difference in the killing between BT10432 and E264 in Madagascar hissing cockroach infection models. The Paenibacilli display varying degrees of pathogenicity, with *P. apiarius* being the most pathogenic of all the bacteria, killing within 48 hrs. *P. thiaminolyticus* is more pathogenic than *P. dendritiformis*.

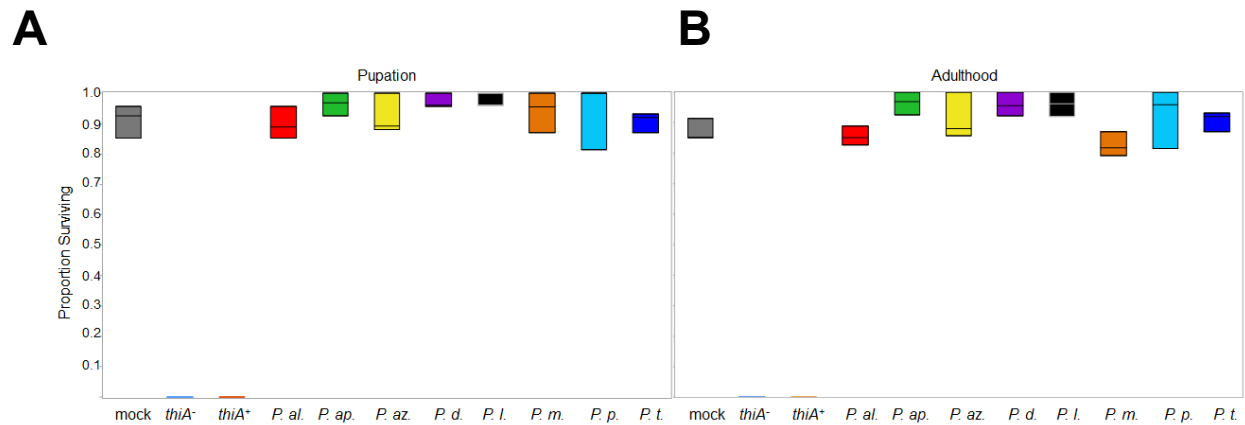


Figure 4 Larval infections with thiaminase I and non-thiaminase I *B. thailandensis* and *Paenibacillus* strains. (A) demonstrates the survival to pupation after infection at the L2 stage. Both *B. thailandensis* strains are pathogenic, demonstrating that ThiA is not important for the pathogenicity. (B) shows the survival to adulthood. There is no significant difference between any *Paenibacilli* and the mock-infected control. *thiA*⁻ = BT10432, *thiA*⁺ = E264, *P. al.* = *P. alvei*, *P. ap.* = *P. apiarius*, *P. az.* = *P. azotofixans*, *P. d.* = *P. dendritiformis*, *P. l.* = *P. larvae pulvifaciens*, *P. m.* = *P. macerans*, *P. p.* = *P. polymyxa*, and *P. t.* = *P. thiaminolyticus*.

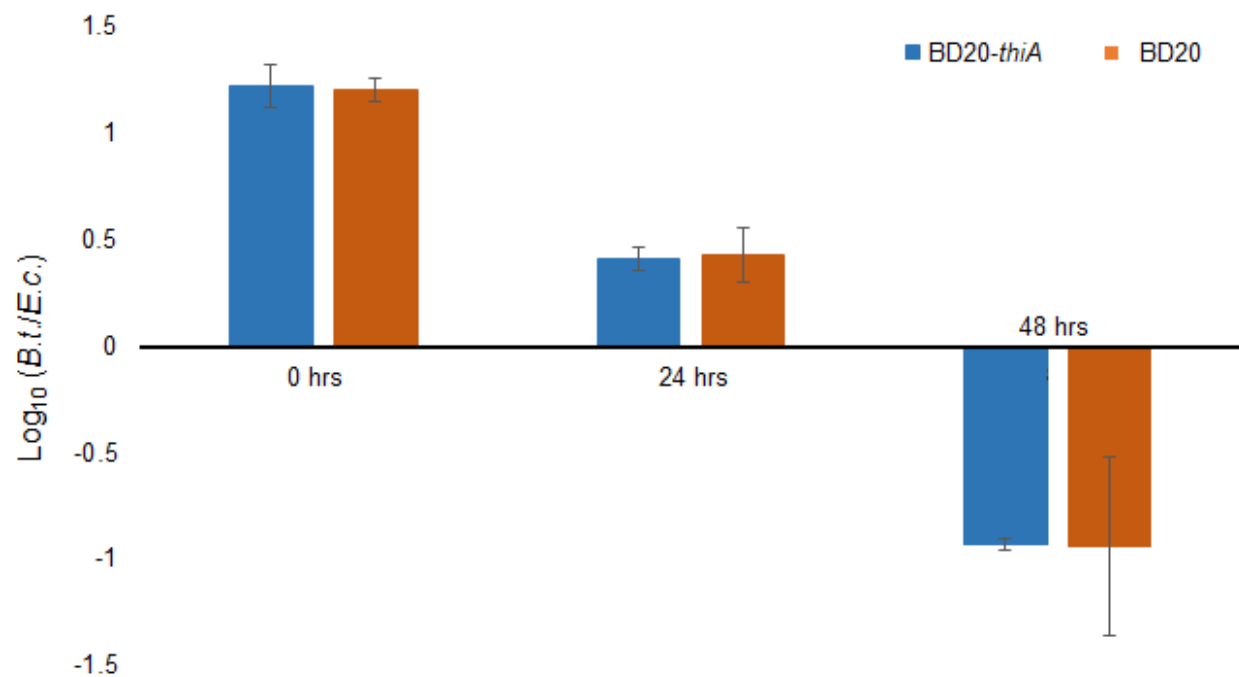


Figure 5 Head to head infections of $\Delta btaK$ *B. thailandensis* strains and a thiamin auxotrophic *E. coli* strain. The graph shows the ratio of *B. thailandensis* to *E. coli* at 0 hrs, 24 hrs, and 48 hrs with \pm standard error bars. At each time point there is no difference between the BD20-*thiA* and BD20 strains, and *E. coli* actually appears to be outcompeting the *B. thailandensis* strains.

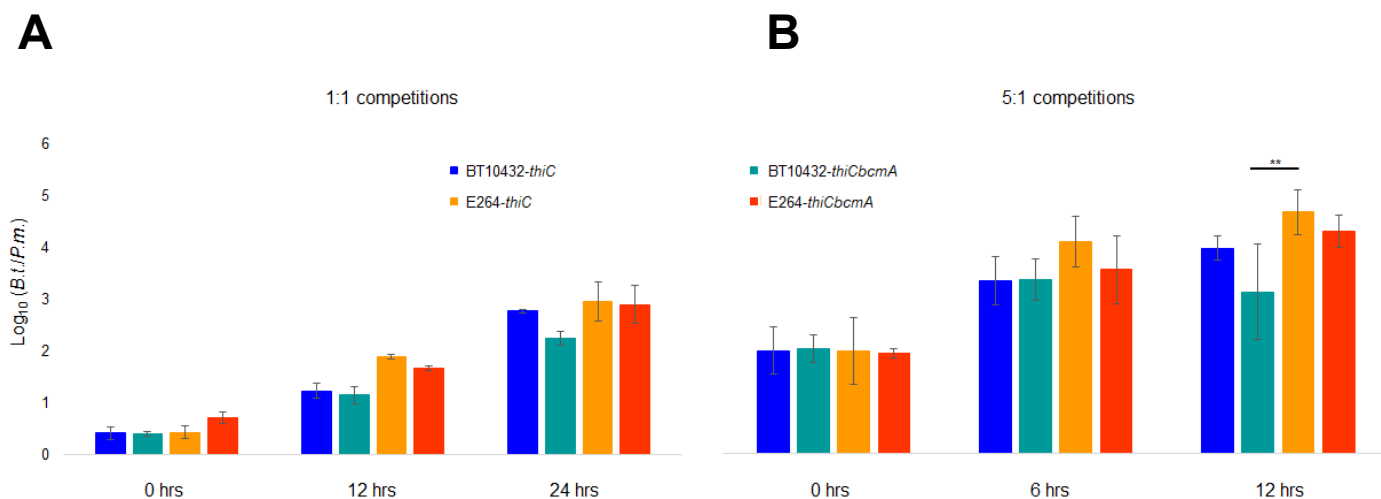


Figure 6 ThiA sensitizes *P. macerans* to bacimethrin in head to head competitions against *B. thailandensis*. (A) shows the head to head competition at a 1:1 ratio of *B. thailandensis* strains to *P. macerans*. Although there were no significant competition effects, the ratio of *B. thailandensis* to *P. macerans* in the BT10432-*thiCbcmA* competitions were lower than all other competitions, and the E264-*thiC* tended to have the highest ratio. (B) demonstrates the head to head competition at a 5:1 ratio of *B. thailandensis* strains to *P. macerans*. At 12 hrs, we observed a significant difference, when comparing between all the samples at each hour set, between the ratio of *B. thailandensis* to *P. macerans* in the BT10432-*thiC* and E264-*thiC* competitions ($p = 0.0078$ after Bonferroni correction). The E264-*thiC* had a higher ratio of *B. thailandensis*, suggesting that both ThiA and bacimethrin are important for a full competitive advantage.

Table 2 Growth of challengers in $\Delta btaK$ *B. thailandensis* strain treated media.

Challenger	12hr BT10432- <i>thiCthiG</i>	24hr BT10432- <i>thiCthiG</i>	36hr BT10432- <i>thiCthiG</i>	48hr BT10432- <i>thiCthiG</i>	60hr BT10432- <i>thiCthiG</i>
<i>A. viridans</i>	Y	Y	Y	Y	Y
<i>L. plantarum</i>	Y	Y	Y	Y	Y
<i>P. alvei</i>	Y	Y	Y	Y	Y
<i>P. macerans</i>	Y	Y	Y	Y	Y
Challenger	12hr E264- <i>thiCthiG</i>	24hr E264- <i>thiCthiG</i>	36hr E264- <i>thiCthiG</i>	48hr E264- <i>thiCthiG</i>	60hr E264- <i>thiCthiG</i>
<i>A. viridans</i>	Y	Y	Y	Y	Y
<i>L. plantarum</i>	Y	Y	Y	Y	Y
<i>P. alvei</i>	Y	Y	Y	Y	Y
<i>P. macerans</i>	Y	Y	Y	Y	Y

Table 3 Growth of challengers in *thiC-thiG*- *B. thailandensis* strain treated media

Challenger	12hr BT10432- <i>thiCthiG</i>	24hr BT10432- <i>thiCthiG</i>	36hr BT10432- <i>thiCthiG</i>	48hr BT10432- <i>thiCthiG</i>	60hr BT10432- <i>thiCthiG</i>
<i>A. viridans</i>	Y	N	N	N	N
<i>L. plantarum</i>	Y	Y	Y	N	N
<i>P. alvei</i>	Y	Y	Y	Y	Y
<i>P. macerans</i>	Y	N	N	N	N
Challenger	12hr E264- <i>thiCthiG</i>	24hr E264- <i>thiCthiG</i>	36hr E264- <i>thiCthiG</i>	48hr E264- <i>thiCthiG</i>	60hr E264- <i>thiCthiG</i>
<i>A. viridans</i>	Y	N	N	N	N
<i>L. plantarum</i>	Y	Y	Y	N	N
<i>P. alvei</i>	Y	Y	Y	Y	Y
<i>P. macerans</i>	N	N	N	N	N

Table 4 Growth of challengers in purified thiaminase I treated media

Challenger	TSB	TSB + BcmE	DM4	DM4 + 1mM thiamin	DM4 + BcmE
<i>A. viridans</i>	Y	Y	N	N	N
<i>L. plantarum</i>	Y	Y	Y	Y	Y
<i>P. alvei</i>	Y	Y	poor growth	poor growth	poor growth
<i>P. macerans</i>	Y	Y	Y	Y	Y

A.4 DISCUSSION

Our results did not clearly establish thiaminase I as a pathogenicity factor in *thiA* containing bacteria. This could be due to a number of reasons. In the case of *B. thailandensis*, it has a robust system for interacting with animals (27). The genome encodes for 5 type VI secretion systems (T6SS), one of which is necessary for pathogenicity in mice (27), and the type III secretion system was also shown to be necessary for intracellular replication as the cells are unable to escape the intracellular vacuole of mammalian cells during infection (26). These 2 systems were shown to be unnecessary for infection in *D. melanogaster* (29), while it was found that *B. thailandensis* that do not have a T6SS-1 were avirulent in Madagascar hissing cockroaches (28), suggesting a role in evading innate immunity. The lack of a shared factor necessary for infecting both insect species suggests that *B. thailandensis* is equipped with other,

uncharacterized factors to evade host immunity and infect and kill insects. They encode for polyketide synthases and non-ribosomal peptide synthases that produce compounds like malleilactone which is necessary for pathogenicity in *C. elegans* and *Dictyostelium discoideum* (65). The lack of a ThiA response in both *D. melanogaster* and Madagascar hissing cockroaches suggests ThiA is not one of these important factors for pathogenesis. Investigations into the proteome of *B. thailandensis* shows that of the 368 known virulence genes found in *B. pseudomallei*, *B. thailandensis* has orthologs of 275 genes, with greater than 80% average similarity (66). The diversity in the ability to make virulence factors may make it too difficult to pin down pathogenicity to ThiA. Only when we infected flies with *B. thailandensis* thiamin auxotrophs, a slight ThiA effect was observed, but it was barely significant. It is possible that we would observe a more significant effect when infecting thiamin deficient flies or flies reared on a low thiamin diet. In these cases there might be less free thiamin and thiamin precursors available for *B. thailandensis* to acquire inside the host, thus having a ThiA would give it an advantage. The thiamin available would be converted to the precursors, which it appears to prefer to grow on, both limiting available thiamin for the host and supplementing its thiamin auxotrophy. The BT10432-*thiC* mutant would not be able to do this and could potentially be at a disadvantage. Further, in another study, we discovered that the microbiota contributes thiamin to the host, thus infections of axenic flies on lower thiamin diets may show a ThiA effect. *B. thailandensis* is capable of infecting insects, it could also be that there is no observable ThiA effect because we are not providing the right insect host for these infection experiments. Another possibility is that the starting OD₆₀₀ were too high, and at a low initial concentration of infectious cells could show a possible effect, as there may be too many cells present for infection to observe a real ThiA promoted advantage.

Due to our lack of a true negative control, it is impossible to ascribe the pathogenicity observed in the *Paenibacillus thiA* species to ThiA production. The fact that all 3 are pathogenic does suggest there is a correlation between having a thiaminase I and being infectious. The ability to produce thiamin may have an effect on pathogenicity as well as *P. apiarius* is thiamin capable and kills the fastest, while the other 2 are auxotrophs with differences in killing speed. Both have the same genomic capabilities for salvaging thiamin, so it may be that the *P. thiaminolyticus* thiaminase I is more active inside the host, and can acquire more precursors for thiamin salvage. Nano-injections of pure ThiA into the hemolymph of *D. melanogaster* would definitively show if ThiA alone could cause pathogenicity in *D. melanogaster*. Also observing the growth of the bacteria within the host may help further elucidate how ThiA contributes to pathogenicity. Experiments were attempted, however, they failed for various reasons (time 0's did not grow, contamination issues, etc.), these experiments need to be repeated to understand this.

Although more difficult to work with, the Madagascar hissing cockroaches may offer the ability to perform experiments not possible in *D. melanogaster*. Due to its size, it might be possible to perform thiamin assays on the hemolymph prior to infection, and then after infection, if the sampling is not too invasive. This would provide information if ThiA is depleting thiamin levels within the hemolymph. It may also allow for real time tracking of bacterial progression within the hemolymph. Flies are too small and need to be homogenized completely, but it may be possible to remove a small volume of hemolymph throughout an infection cycle and plate it to see the progression within the same organism. It might be possible to achieve non-lethal hemolymph removal through a nano-injector. The problems with the Madagascar hissing cockroaches is their housing and numbers. They lack the replicative power that *D. melanogaster*

has, but using both systems in conjunction with one another may help further elucidate ThiA's potential role in pathogenicity.

Paenibacillus species include known insect pathogens as well as species that can produce insecticides (56), making it possible that the ThiA producers are already geared towards insect pathogenicity, even though they currently are not recognized as insect pathogens. We recently sequenced the genomes of *P. apiarius* NRRL B-23460 and *P. thiaminolyticus* NRRL B-4156, and when investigating these genomes, as well as the genome of *P. dendritiformis* C454, genomic evidence suggests these organisms are geared for insect pathogenicity. This becomes clearer when comparing their genomic capabilities to the pathogen *P. alvei* DSM 29. Chitin, a polymer of N-acetyl- β -D-glucosamine is essential to insects as it is a major component of their exoskeleton as well as their peritrophic matrix, which serves as their gut barrier (67). The genome of *P. alvei* DSM 29 encodes for multiple chitinases, allowing for the potential to break down insect chitin (68). The genomes of *P. apiarius*, *P. dendritiformis*, and *P. thiaminolyticus* encode for multiple chitinases and chitin binding proteins, with the *P. thiaminolyticus* genome also encoding for a chitosanase. As in *P. alvei* DSM 29, the thiaminase I producing *Paenibacilli* genomes encode for a hyaluronate lyase, an enzyme that degrades hyaluron, a polysaccharide found in animal connective tissues (69). This enzyme is produced by *Streptococcus agalactiae* during infection, and higher production of hyaluronate lyase corresponds with higher virulence (69). The *Paenibacilli* may be able to degrade the connective tissue of dead insects and other animals with this enzyme, and this along with the chitinases may also allow for them to infect insects through an external route.

Alveolysin is a thiol-activated toxin with homology to listeriolysin o and other putative virulence factors that lyse membranes and sequester cholesterol, and it is also encoded for in the

P. alvei genome (68, 70). Alveolysin is present in both *P. dendritiformis* and *thiaminolyticus*, but is absent in the *P. apiarius* genome. Mosquitocidal toxin (MTX) is an ADP-ribosylating protein toxic to mosquito larvae (71), and it is present in the *P. alvei* genome. *P. apiarius* has a homolog of it present in its genome, while no homologs of the *P. alvei* MTX were detected in *P. dendritiformis* and *P. thiaminolyticus*. However, both *P. dendritiformis* and *thiaminolyticus* have an MTX2 like protein in their genomes, which may also serve as an insecticidal toxin. Unlike *P. apiarius*, the genomes of *P. dendritiformis* and *thiaminolyticus* encode for a binary toxin system. Both genomes contain two protective antigen proteins followed by a lethal factor domain protein and an ADP ribosyltransferase protein. Both the lethal factor protein and the ADP ribosyltransferase proteins contain vegetative insecticidal protein (VIP2) domains which are found in toxins that ribosylate actin. These genomic factors may indicate that all three *Paenibacilli* may be more pathogenic towards insects than originally identified. *P. dendritiformis* and *P. thiaminolyticus* appear to be especially geared towards insect pathogenesis with the presence of alveolysin and a binary toxin system with a lethal factor domain and ADP-ribosylation protein geared for insect cells. All three strains are lacking the orally active insecticidal toxin complex found in *P. alvei* that is homologous to the complex found in *Bacillus thuringiensis* IBL 200 (68). These may all account for the pathogenicity experienced during the infections.

All our attempts at generating *thiA*⁻ *Paenibacilli* failed, and it is unclear why this occurred. We utilized multiple antibiotic resistance markers after screening the species' abilities to grow, in hopes that one would be good for the transformation. We utilized different types of DNA (circular and linearized plasmids and PCR products, as well as denatured plasmids) in hopes that one would work. It is unclear if DNA was able to enter the cells through the 3

electroporation methods we used or if the problem was that the DNA was unable to recombine after it entered the cell. *Paenibacillus* species are known to be difficult to transform, so methods need to be applied to generate a working protocol. Recently, Bach and colleagues optimized a protocol for transforming the recalcitrant *P. riograndensis* SBR5^T (72). In this study, they investigated multiple aspects of the transformation protocol including which growth mediums to use, when to collect cultures for generating electrocompetent cells, the effectiveness of cell wall-disturbing agents, proper washing steps, proper recovery media to use, and how much DNA to add. This comprehensive study found specific media (BHI, BHIS, KB, and MYPGP) better for getting transformants, and found that harvesting cells at the OD₆₀₀ of 0.3 gave the peak amount of transformants (72). Further they found that certain media were better for using for recovery than others, and that 100 ng of DNA was the best concentration. They also found that using *E. coli* ER2925 to replicate the plasmid (this strain does not produce an *E. coli* methylation pattern), led to higher transformants. Applying these optimization techniques to the 3 *thiA* species may allow for transformation to occur. If it is possible to at the very least get a plasmid to replicate within the cells through an efficient electroporation protocol, but not homologous recombination, CRISPR could be used to generate *thiA*⁻ strains.

The lack of pathogenicity in an oral infection route was somewhat surprising, especially for the *B. thailandensis* strains, as E264 was shown to be orally infective (29). There are some caveats to this, first the fly food used was different, containing fructose, dry milk powder, and dehydrated mashed potatoes in a dry mix. This form of oral infection may actually disrupt the peritrophic matrix and lead to systemic infection, leading to the death observed, accounting for why we did not observe any death. Feeding experiments should be conducted on completely chemically defined diets varying in thiamin concentration using E264 and BT10432 background

mutants. It may be that in flies that have developed on a low thiamin diet would be more susceptible to infection due to the presence of ThiA. To mitigate the effects of microbiota derived thiamin, this could also be conducted on axenic flies to directly see if the interaction between thiaminase I producers and diet influences host health. To see if there is a ThiA effect in larval infections, again a completely chemically defined fly diet could be used with the *Paenibacilli*.

Based on our findings, in our experiments thiaminase I alone does not appear to be an effective competition factor. Purified thiaminase I did not prevent growth of any of the known auxotrophs and may have actually promoted the growth of *P. macerans*. This could be due to production of precursors that *P. macerans* and the other 3 auxotrophs can potentially utilize. All 4 species were unable to make HMP based on their genomic data, suggesting that they are able to utilize the modified HMP generated by thiaminase I, seemingly refuting the hypothesis that this compound would be toxic and a competition factor within itself. *B. thailandensis* was only able to limit competitor growth when it was capable of producing bactobolin A, suggesting that *B. thailandensis* is competitive against other bacteria using an arsenal of different competition factors (27, 48). Schwarz and colleagues found that their T6SS-1 is responsible for cell-to-cell contact against certain proteobacteria, but in competitions with Firmicutes it was competitive against them in a T6SS-1 independent fashion, suggesting it has other mechanisms for out-competing them (27). Possibly products from their polyketide and non-ribosomal protein synthases are involved in bacterial interactions, and thiaminase I may be one small part of their competition system.

ThiA does appear to play a role in sensitizing *P. macerans* as there is a significantly lower ratio of *B. thailandensis* to *P. macerans* when BT10432-*thiCbcmA* is competed in contrast

to when E264-*thiC* is competed. There is no significant difference when E264-*thiCbcmA* is competed suggesting that bacimethrin synthesis is necessary for thiaminase I to be effective. The degradation of thiamin possibly leads to the increase in transport of bacimethrin by the competitor as it is reliant on external sources for either salvaging thiamin or its precursors. When thiamin is not present in the surrounding media due to ThiA degradation, *P. macerans* is likely to express transporters for HMP for salvage. The production of bacimethrin by *B. thailandensis* would lead to *P. macerans* importing it instead of HMP, and dying as a result. To really understand if ThiA is sensitizing competitors to bacimethrin, experiments should be conducted with both purified thiaminase I and bacimethrin. Our experiments show purified ThiA added back to media is ineffective at preventing the growth of other cells, but we do not know what happens when purified bacimethrin is added to media, or what happens if it is added in conjunction with thiaminase I. Doing this in a defined media would be better as it would allow for certain manipulations. If HMP production by ThiA is responsible for the uptake of bacimethrin, experiments can be set up to see the degree of antimicrobial activity of bacimethrin when thiamin or HMP is present, as well as when either of those 2 compounds and thiaminase I is present. For clearer competition results, generation of a chemically defined media in which both *B. thailandensis* and the competitor grow at similar rates would be necessary. It is possible that in the head to head competitions, due to differences in growth (*P. macerans* grows faster in TSB than *B. thailandensis*), the conditions might not be right for accurate competition. This is likely the case with the *E. coli* competitions as they grew much faster than *B. thailandensis*. Further using a completely defined media would allow us to control the amount of thiamin and precursors present in the media.

A.5 REFERENCES

1. Kraft CE, Angert ER. 2017. Competition for vitamin B1 (thiamin) structures numerous ecological interactions. *The Quarterly Review of Biology* 92:151-168.
2. Jurgenson CT, Begley TP, Ealick SE. 2009. The structural and biochemical foundations of thiamin biosynthesis. *Annu Rev Biochem* 78:569-603.
3. Schauer K, Stolz J, Scherer S, Fuchs TM. 2009. Both thiamine uptake and biosynthesis of thiamine precursors are required for intracellular replication of *Listeria monocytogenes*. *Journal of bacteriology* 191:2218-2227.
4. Webb E, Claas K, Downs D. 1998. thiBPQ Encodes an ABC Transporter Required for Transport of Thiamine and Thiamine Pyrophosphate in *Salmonella typhimurium*. *Journal of Biological Chemistry* 273:8946-8950.
5. Karunakaran R, Ebert K, Harvey S, Leonard M, Ramachandran V, Poole P. 2006. Thiamine is synthesized by a salvage pathway in *Rhizobium leguminosarum* bv. *viciae* strain 3841. *Journal of bacteriology* 188:6661-6668.
6. Carini P, Campbell EO, Morré J, Sanudo-Wilhelmy SA, Thrash JC, Bennett SE, Temperton B, Begley T, Giovannoni SJ. 2014. Discovery of a SAR11 growth requirement for thiamin's pyrimidine precursor and its distribution in the Sargasso Sea. *ISME J* 8:1727-1738.
7. Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS. 2002. Comparative genomics of thiamin biosynthesis in prokaryotes New genes and regulatory mechanisms. *Journal of Biological chemistry* 277:48949-48959.
8. Rodionov DA, Hebbeln P, Eudes A, Ter Beek J, Rodionova IA, Erkens GB, Slotboom DJ, Gelfand MS, Osterman AL, Hanson AD. 2009. A novel class of modular transporters for vitamins in prokaryotes. *Journal of bacteriology* 191:42-51.
9. Devedjiev Y, Surendranath Y, Derewenda U, Gabrys A, Cooper DR, Zhang R-g, Lezondra L, Joachimiak A, Derewenda ZS. 2004. The structure and ligand binding properties of the *B. subtilis* YkoF gene product, a member of a novel family of thiamin/HMP-binding proteins. *Journal of molecular biology* 343:395-406.
10. Anderson LN, Koech PK, Plymale AE, Landorf EV, Konopka A, Collart FR, Lipton MS, Romine MF, Wright AT. 2015. Live cell discovery of microbial vitamin transport and enzyme-cofactor interactions. *ACS chemical biology* 11:345-354.
11. Toms AV, Haas AL, Park J-H, Begley TP, Ealick SE. 2005. Structural characterization of the regulatory proteins TenA and TenI from *Bacillus subtilis* and identification of TenA as a thiaminase II. *Biochemistry* 44:2319-2329.
12. Benach J, Edstrom WC, Lee I, Das K, Cooper B, Xiao R, Liu J, Rost B, Acton TB, Montelione GT. 2005. The 2.35 Å structure of the TenA homolog from *Pyrococcus furiosus* supports an enzymatic function in thiamine metabolism. *Acta Crystallographica Section D: Biological Crystallography* 61:589-598.
13. Zallot R, Yazdani M, Goyer A, Ziemak MJ, Guan J-C, McCarty DR, de Crécy-Lagard V, Gerdes S, Garrett TJ, Benach J. 2014. Salvage of the thiamin pyrimidine moiety by plant TenA proteins lacking an active-site cysteine. *Biochemical Journal* 463:145-155.
14. Onozuka M, Konno H, Kawasaki Y, Akaji K, Nosaka K. 2007. Involvement of thiaminase II encoded by the THI20 gene in thiamin salvage of *Saccharomyces cerevisiae*. *FEMS yeast research* 8:266-275.

15. Jenkins AH, Schyns G, Potot S, Sun G, Begley TP. 2007. A new thiamin salvage pathway. *Nature chemical biology* 3:492-497.
16. Costello CA, Kelleher NL, Abe M, McLafferty FW, Begley TP. 1996. Mechanistic studies on thiaminase I overexpression and identification of the active site nucleophile. *Journal of Biological Chemistry* 271:3445-3452.
17. Fujita A. 1954. Thiaminase. *Advances in Enzymology and Related Areas of Molecular Biology*, Volume 15:389-421.
18. Agee CC, Airth R. 1973. Reversible inactivation of thiaminase I of *Bacillus thiaminolyticus* by its primary substrate, thiamine. *Journal of bacteriology* 115:957-965.
19. Ebata J, Murata K. 1961. The purification of thiaminase I produced by *Bacillus thiaminolyticus*. *The Journal of vitaminology* 7:115-121.
20. Kreinbring CA, Remillard SP, Hubbard P, Brodtkin HR, Leeper FJ, Hawksley D, Lai EY, Fulton C, Petsko GA, Ringe D. 2014. Structure of a eukaryotic thiaminase I. *Proceedings of the National Academy of Sciences* 111:137-142.
21. Shreeve J, Edwin E. 1974. Thiaminase-producing strains of *Cl. Sporogenes* associated with outbreaks of cerebrocortical necrosis. *The Veterinary record* 94:330.
22. Honeyfield DC, Hinterkopf JP, Brown SB. 2002. Isolation of thiaminase-positive bacteria from alewife. *Transactions of the American Fisheries Society* 131:171-175.
23. Richter CA, Evans AN, Wright-Osment MK, Zajicek JL, Heppell SA, Riley SC, Krueger CC, Tillitt DE. 2012. *Paenibacillus thiaminolyticus* is not the cause of thiamine deficiency impeding lake trout (*Salvelinus namaycush*) recruitment in the Great Lakes. *Canadian Journal of Fisheries and Aquatic Sciences* 69:1056-1064.
24. Kuno Y. 1951. *Bacillus thiaminolyticus*, a new thiamin-decomposing bacterium. *Proceedings of the Japan Academy* 27:362-365.
25. Ringe H, Schuelke M, Weber S, Dorner BG, Kirchner S, Dorner MB. 2014. Infant botulism: is there an association with thiamine deficiency? *Pediatrics* 134:e1436-e1440.
26. Haraga A, West TE, Brittnacher MJ, Skerrett SJ, Miller SI. 2008. *Burkholderia thailandensis* as a model system for the study of the virulence-associated type III secretion system of *Burkholderia pseudomallei*. *Infection and immunity* 76:5402-5411.
27. Schwarz S, West TE, Boyer F, Chiang W-C, Carl MA, Hood RD, Rohmer L, Tolker-Nielsen T, Skerrett SJ, Mougous JD. 2010. *Burkholderia* type VI secretion systems have distinct roles in eukaryotic and bacterial cell interactions. *PLoS pathogens* 6:e1001068.
28. Fisher NA, Ribot WJ, Applefeld W, DeShazer D. 2012. The Madagascar hissing cockroach as a novel surrogate host for *Burkholderia pseudomallei*, *B. mallei* and *B. thailandensis*. *BMC microbiology* 12:117.
29. Pilátová M, Dionne MS. 2012. *Burkholderia thailandensis* is virulent in *Drosophila melanogaster*. *PLoS One* 7:e49745.
30. Nakamura L. 1990. *Bacillus thiaminolyticus* sp. nov., nom. rev. *International Journal of Systematic and Evolutionary Microbiology* 40:242-246.
31. Katznelson H. 1955. *Bacillus apiarius*, n. sp., an aerobic spore-forming organism isolated from honeybee larvae. *Journal of bacteriology* 70:635.
32. Haughton BG, King H. 1958. Toxypyrimidine phosphate as an inhibitor of bacterial enzyme systems that require pyridoxal phosphate. *Biochemical Journal* 70:660.
33. Cooper LE, O'Leary SnE, Begley TP. 2014. Biosynthesis of a Thiamin Antivitamin in *Clostridium botulinum*. *Biochemistry* 53:2215-2217.

34. Reddick JJ, Saha S, Lee J-m, Melnick JS, Perkins J, Begley TP. 2001. The mechanism of action of bacimethrin, a naturally occurring thiamin antimetabolite. *Bioorganic & medicinal chemistry letters* 11:2245-2248.
35. Broderick NA, Buchon N, Lemaitre B. 2014. Microbiota-induced changes in *Drosophila melanogaster* host gene expression and gut morphology. *MBio* 5:e01117-14.
36. Lemaitre B, Reichhart J-M, Hoffmann JA. 1997. *Drosophila* host defense: differential induction of antimicrobial peptide genes after infection by various classes of microorganisms. *Proceedings of the National Academy of Sciences* 94:14614-14619.
37. Kraft CE, Gordon ER, Angert ER. 2014. A Rapid Method for Assaying Thiaminase I Activity in Diverse Biological Samples. *PloS one* 9:e92688.
38. Buchon N, Broderick NA, Kuraishi T, Lemaitre B. 2010. *Drosophila* EGFR pathway coordinates stem cell proliferation and gut remodeling following infection. *BMC biology* 8:152.
39. Shanks RM, Caiazza NC, Hinsa SM, Toutain CM, O'Toole GA. 2006. *Saccharomyces cerevisiae*-based molecular tool kit for manipulation of genes from gram-negative bacteria. *Applied and environmental microbiology* 72:5027-5036.
40. Shanks RM, Kadouri DE, MacEachran DP, O'Toole GA. 2009. New yeast recombineering tools for bacteria. *Plasmid* 62:88-97.
41. Gallagher LA, Ramage E, Patrapuvich R, Weiss E, Brittnacher M, Manoil C. 2013. Sequence-defined transposon mutant library of *Burkholderia thailandensis*. *MBio* 4:e00604-13.
42. Basset A, Tzou P, Lemaitre B, Boccard F. 2003. A single gene that promotes interaction of a phytopathogenic bacterium with its insect vector, *Drosophila melanogaster*. *EMBO reports* 4:205-209.
43. Peters JE. 2007. Gene transfer in Gram-negative bacteria, p 735-755, *Methods for General and Molecular Microbiology*, Third Edition. American Society of Microbiology.
44. Zarschler K, Janesch B, Zayni S, Schäffer C, Messner P. 2009. Construction of a gene knockout system for application in *Paenibacillus alvei* CCM 2051T, exemplified by the S-layer glycan biosynthesis initiation enzyme WsfP. *Applied and environmental microbiology* 75:3077-3085.
45. Murray KD, Aronstein KA. 2008. Transformation of the Gram-positive honey bee pathogen, *Paenibacillus* larvae, by electroporation. *Journal of microbiological methods* 75:325-328.
46. Kim S-B, Timmusk S. 2013. A simplified method for gene knockout and direct screening of recombinant clones for application in *Paenibacillus polymyxa*. *PloS one* 8:e68092.
47. Thongdee M, Gallagher LA, Schell M, Dharakul T, Songsivilai S, Manoil C. 2008. Targeted mutagenesis of *Burkholderia thailandensis* and *Burkholderia pseudomallei* through natural transformation of PCR fragments. *Applied and environmental microbiology* 74:2985-2989.
48. Chandler JR, Heilmann S, Mittler JE, Greenberg EP. 2012. Acyl-homoserine lactone-dependent eavesdropping promotes competition in a laboratory co-culture model. *The ISME journal* 6:2219.
49. Newell PD, Chaston JM, Wang Y, Winans NJ, Sannino DR, Wong AC, Dobson AJ, Kagle J, Douglas AE. 2014. In vivo function and comparative genomic analyses of the *Drosophila* gut microbiota identify candidate symbiosis factors. *Front Microbiol* 5:576.

50. Zhang K, Bian J, Deng Y, Smith A, Nunez RE, Li MB, Pal U, Yu A-M, Qiu W, Ealick SE. 2016. Lyme disease spirochaete *Borrelia burgdorferi* does not require thiamin. *Nature microbiology* 2:16213.
51. Nakamura L. 1996. *Paenibacillus apiarius* sp. nov. *International Journal of Systematic and Evolutionary Microbiology* 46:688-693.
52. Djordjevic SP, Forbes WA, Smith LA, Hornitzky MA. 2000. Genetic and Biochemical Diversity among Isolates of *Paenibacillus alvei* Cultured from Australian Honeybee (*Apis mellifera*) Colonies. *Applied and environmental microbiology* 66:1098-1106.
53. Heyndrickx M, Vandemeulebroecke K, Hoste B, Janssen P, Kersters K, De Vos P, Logan N, Ali N, Berkeley R. 1996. Reclassification of *Paenibacillus* (formerly *Bacillus*) *pulvifaciens* (Nakamura 1984) Ash et al. 1994, a later subjective synonym of *Paenibacillus* (formerly *Bacillus*) *larvae* (White 1906) Ash et al. 1994, as a subspecies of *P. larvae*, with emended descriptions of *P. larvae* as *P. larvae* subsp. *larvae* and *P. larvae* subsp. *pulvifaciens*. *International Journal of Systematic and Evolutionary Microbiology* 46:270-279.
54. Dingman DW. 2009. DNA fingerprinting of *Paenibacillus popilliae* and *Paenibacillus lentimorbus* using PCR-amplified 16S–23S rDNA intergenic transcribed spacer (ITS) regions. *Journal of invertebrate pathology* 100:16-21.
55. Rosado AS, Duarte GF, Seldin L, Van Elsas JD. 1998. Genetic Diversity of *nifH* Gene Sequences in *Paenibacillus azotofixans* Strains and Soil Samples Analyzed by Denaturing Gradient Gel Electrophoresis of PCR-Amplified Gene Fragments. *Applied and Environmental Microbiology* 64:2770-2779.
56. Grady EN, MacDonald J, Liu L, Richman A, Yuan Z-C. 2016. Current knowledge and perspectives of *Paenibacillus*: a review. *Microbial cell factories* 15:203.
57. Timmusk S, Wagner EGH. 1999. The plant-growth-promoting rhizobacterium *Paenibacillus polymyxa* induces changes in *Arabidopsis thaliana* gene expression: a possible connection between biotic and abiotic stress responses. *Molecular Plant-Microbe Interactions* 12:951-959.
58. Lemaitre B, Hoffmann J. 2007. The host defense of *Drosophila melanogaster*. *Annu Rev Immunol* 25:697-743.
59. Rutschmann S, Kilinc A, Ferrandon D. 2002. Cutting edge: the toll pathway is required for resistance to gram-positive bacterial infections in *Drosophila*. *The Journal of Immunology* 168:1542-1546.
60. Church RB, Robertson FW. 1966. A biochemical study of the growth of *Drosophila melanogaster*. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology* 162:337-351.
61. Greenlee KJ, Montooth KL, Helm BR. 2014. Predicting performance and plasticity in the development of respiratory structures and metabolic systems. *Integrative and comparative biology* 54:307-322.
62. May CM, Doroszuk A, Zwaan BJ. 2015. The effect of developmental nutrition on life span and fecundity depends on the adult reproductive environment in *Drosophila melanogaster*. *Ecology and evolution* 5:1156-1168.
63. Sang JH. 1956. The quantitative nutritional requirements of *Drosophila melanogaster*. *J Exp Biol* 33:45-72.
64. Sang JH. 1962. Relationships between protein supplies and B-vitamin requirements, in axenically cultured *Drosophila*. *Journal of Nutrition* 77:355-368.

65. Biggins JB, Ternei MA, Brady SF. 2012. Malleilactone, a polyketide synthase-derived virulence factor encoded by the cryptic secondary metabolome of *Burkholderia pseudomallei* group pathogens. *Journal of the American Chemical Society* 134:13192-13195.
66. Yu Y, Kim HS, Chua HH, Lin CH, Sim SH, Lin D, Derr A, Engels R, DeShazer D, Birren B. 2006. Genomic patterns of pathogen evolution revealed by comparison of *Burkholderia pseudomallei*, the causative agent of melioidosis, to avirulent *Burkholderia thailandensis*. *BMC microbiology* 6:46.
67. Merzendorfer H, Zimoch L. 2003. Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. *Journal of Experimental Biology* 206:4393-4412.
68. Djukic M, Becker D, Poehlein A, Voget S, Daniel R. 2012. Genome sequence of *Paenibacillus alvei* DSM 29, a secondary invader during European foulbrood outbreaks. *Journal of bacteriology* 194:6365-6365.
69. Li S, Jedrzejewski MJ. 2001. Hyaluronan Binding and Degradation by *Streptococcus agalactiae* Hyaluronate Lyase. *Journal of Biological Chemistry* 276:41407-41416.
70. Geoffroy C, Mengaud J, Alouf J, Cossart P. 1990. Alveolysin, the thiol-activated toxin of *Bacillus alvei*, is homologous to listeriolysin O, perfringolysin O, pneumolysin, and streptolysin O and contains a single cysteine. *Journal of bacteriology* 172:7301-7305.
71. Carpusa I, Jank T, Aktories K. 2006. *Bacillus sphaericus* mosquito-cidal toxin (MTX) and pierisin: the enigmatic offspring from the family of ADP-ribosyltransferases. *Molecular microbiology* 62:621-630.
72. Bach E, de Carvalho Fernandes G, Passaglia LMP. 2016. How to transform a recalcitrant *Paenibacillus* strain: From culture medium to restriction barrier. *Journal of microbiological methods* 131:135-143.

APPENDIX B

ATTEMPTS AT CULTURING *EPULOPISCIMUM* SPP. AND THEIR RELATIVES

B.1 INTRODUCTION

Epulopiscium spp. are a group of bacteria (epulos for short) that live as intestinal symbionts in the guts of marine tropical and herbivorous and detritivorous surgeonfish from the family Acanthuridae. These bacteria were originally thought to be protists based on their size and were categorized into different morphotypes based on their size and morphology (1). Through 16S rRNA analysis, it was determined that these organisms were actually bacteria found in the Lachnospiraceae family, the XIVb cluster of the clostridiales (2). These bacteria are extremely unique in the bacterial world as members of this group include the largest known heterotrophic bacteria, as the largest morphotype (type A) exceeds lengths of 600µm and widths of 80µm (3, 4). Epulos have varying degrees of morphologies, multiple morphotypes are cigar shaped (type A1, A2, C, E, H), while others are more ovoid shaped (type B), or long and filamentous (type J) (1, 5). They also display a unique variety of reproductive strategies. Some morphotypes reproduce through binary fission (type E, G, I, and J), however, other morphotypes appear to have completely lost the ability to undergo binary fission. Type A1, A2, and B reproduce through the novel production of live, intracellular offspring, with up to 12 intracellular offspring being observed in type B. This is a derived form of endosporulation (6-8), and the type B (*Ca. Epulopiscium viviparous*) genome contains multiple homologs of genes corresponding to the early and late stages of sporulation found in *Bacillus subtilis*, as well as the physiological hallmarks of the early stages of sporulation (6, 7, 9). Genes present in the genome include the endosporulation specific sigma factors, *spo0A* the master regulator of the initiation of sporulation, spore cortex, and spore coat genes (7). Type C and J form intracellular offspring in

the form of true phase-bright endospores with a dipicolinic acid core. These cells have the ultrastructure of endospores found in *Bacillus* and *Clostridium* species. Types D, E, and F may also produce endospores as well (1).

Epulos are typically found in multiple surgeonfish hosts, and are generally found in mixed populations, though they can be spatially separated within the gut (5, 10). The exception to this is '*Ca. Epulopiscium viviparous*' as it is only found in *Naso tonganus* with no other epulos present. Type A1 and A2s are found together in *Acanthurus lineatus* and *A. nigrofuscus* (11), while type C's and J's are found together in *N. unicornis* and *N. lituratus* (5). Epulo populations appear to be synchronized and grow in diurnal cycles that coincide with host feeding (5, 6, 12). For type C and J epulo populations, during the morning and early day while the fish is active and feeding, the epulos are motile and phase-dark, while undergoing the early stages of offspring development. As the day progresses, the offspring continue to develop within the mother cell, and at night, while the fish is no longer active, the forespores mature into phase-bright endospore. At this point the mother cells are no longer motile and degrade releasing the mature endospores. In the early morning, the endospores germinate into vegetative cells and the cycle perpetuates (5). It is likely that a similar diurnal circadian developmental cycle occurs in populations of other intracellular offspring producing epulos, without complete endosporulation occurring. These factors may make culturing these organisms difficult because there may be environmental or host derived cues that establish this circadian cycle and allow for this complex life cycle. Further it is unknown what triggers epulo endospore germination (13) in endospore forming epulos.

In this study, we attempted to culture epulos from spore stocks from *N. unicornis* and *N. lituratus*, as well as from the feces and gut contents of *Zebrasoma flavescens* maintained in our

lab. Culturing attempts were also carried out in the field on multiple occasions from multiple surgeonfish harboring different epulos. Most attempts at culturing failed, however, we did get some ambiguous and potentially promising results at some stages.

B.2 Culturing attempts at Cornell University

B.3 MATERIALS AND METHODS

Sample collection. Spore samples were collected from *N. unicornis* caught by spearfishing on 4/13/2012 near Coconut Island, Hawaii or from *N. lituratus* and *N. unicornis* caught by spearfishing on 7/6/2013 or purchased from a fish market on 7/13/2016 at and near Coconut Island, Hawaii.

Original media base preparation. The original media base was generated in Hungate flasks. The flasks were filled with N₂ gas for ~20 minutes prior to addition of the media supplements listed in table 1. NaHCO₃ buffer was added to a beaker first and its pH was adjusted to ~7.2. Autoclavable reagents were all mixed in the beaker and it was bubbled with N₂ gas for 20 minutes. The media base was added into the N₂ gassed Hungate flask, which was then capped and sealed. The flask was then autoclaved for 15 minutes. After autoclaving, when the flask was cool enough to be handled, gas exchange was conducted on the headspace for ~20 minutes. While gas exchange was occurring, sterile vitamin mix (Table 2), vitamin K3, and amino acid mix were added to the solution by filter sterilizing them into the media through a 0.2μM filter. Non-autoclavable solutions were prepared by autoclaving capped and sealed Hungate vials that had been filled with N₂ gas, then filter sterilizing the solutions into these vials, and doing gas exchange for ~30 minutes. The influx gas needle was hooked up through a 0.2μM filter attached to a 10mL needle to ensure sterility. Due to initial precipitation problems, CaCl₂ and MgSO₄

were made separately in Hungate vials in 10% solutions (w/v), autoclaved, and the headspace gas was exchanged. These were added directly to the rest of the base after autoclaving. Since supplements were to be added to the basal media, their volumes were accounted for when H₂O was added.

Table 1 original media base

Component	Amount to add for 100mL (in grams)	Amount of stock added for 100mL	Stock solutions % (w/v)
NaCl	0.1		
KCl	0.04		
CaCl₂ 2H₂O		0.2mL	10
NH ₄ NO ₃	0.033		
MgSO₄ 7H₂O		1mL	10
KH ₂ PO ₄	0.02		
K ₂ HPO ₄	0.015		
Casamino acids	0.1		
FeSO ₄	0.00002	100μL	4.00E-02
Na citrate	0.02		
Na pyruvate	0.01		
Hematin		100μL	1.2mg/mL in 200mM L-histidine
Vitamin K3		50μL	1mg/mL in 100% ethanol
vitamin mix		1mL	
trace mineral mix		1mL	
cysteine HCl	0.06		
resazurin		100μL	0.1
Amino acid mix		4.976 mL	1mM
NaHCO ₃	0.5		
H ₂ O		83.284 mL	

Bold font of solutions mean they were not autoclaved in the basal mix, rather added after

Table 2 Vitamin Mix

Vitamin mix	mg/L	To add for 50mL solution	stock % (w/v)	Solubility
pyridoxine HCl	10	500μL	0.1	50mg/mL H ₂ O
riboflavin	5	1.25mL	0.02	10 mg/mL 0.1M NaOH
thiamin HCl	5	50μL	0.5	50mg/mL H ₂ O

Nicotinamide (Niacin)	5	50µL	0.5	50mg/mL H ₂ O
Ca Pantothenate	5	250µL	0.1	50mg/mL H ₂ O
α-Lipoic Acid	5	2.5mL	0.01	70% ethanol
p-aminobenzoic acid	5	125µL	0.2	50mg/mL H ₂ O
folic acid	0.05	200µL	0.05	50mg/mL 1M NH ₄ OH
Biotin	2	1mL	0.01	50mg/mL 1M NaOH
Cobalamin	0.1	100µL	0.005	11.11 mg/mL H ₂ O
KH ₂ PO ₄	900			
H ₂ O		43.975mL		

Table 3 Trace mineral mix

Component	g/L	to add for 50mL solution	Stock % (w/v)
EDTA	0.5	0.025g	
MgSO ₄ 7H ₂ O	3	0.15g	
MnSO ₄ H ₂ O	0.5	0.025g	
NaCl	1	0.05g	
FeSO ₄ 7H ₂ O	0.1	0.005g	
CoCl ₂ 6H ₂ O	0.1	0.005g	
Ca(NO ₃) ₂	0.1	0.005g	
ZnSO ₄ 7H ₂ O	0.1	0.005g	
CuSO ₄ 5H ₂ O	0.01	500µL	0.1
AlK(SO ₄) ₂	0.01	500µL	0.1
H ₃ BO ₃	0.01	500µL	0.1
Na ₂ MoO ₄ 2H ₂ O	0.01	500µL	0.1
Na ₂ SeO ₃	0.001	50µL	0.1
Na ₂ WO ₄ 2H ₂ O	0.01	500µL	0.1
NiCl ₂ 6H ₂ O	0.02	1mL	0.1
H ₂ O		46.3mL	

Table 4 supplements used in first culturing attempts

Component	add for 100mL	Stock % (w/v)	add for 2mL (volume in µL)
D-mannitol	500µL	20	10
D-glucose	500µL	20	10
Fructose	500µL	20	10
D-galactose	1mL	10	20
D-xylose	500µL	20	10
Yeast Extract	1mL	10	20
Tryptone	500µL	20	10
Clarified rumen fluid	200µL		4

Bile salts	1mL	0.1	20
Urea	200μL	0.1	4
Propionic acid	100μL		2
Butyric acid	100μL		2
Acetic acid	85μL		1.7
Isovaleric acid	5μL		0.1
Algal mix	2mL		40

Making supplements. Individual sugars, yeast extract, tryptone, urea, and bile salts were all made as respective stock solutions displayed in Table 4. These were autoclaved for 15 minutes, headspace was exchanged with N₂ for ~20 minutes. To generate the algal extract, dried nori (*Porphyra yezoensis*) and dried brown algae were used. 2.2g of each were ground to a fine powder with a mortar and pestle. H₂O was added to a final volume of 200mL and the pH was dropped to 4 using HCl. Porcine pepsin was then added at 4mg/200mL and incubated at 30°C on a hot plate with constant stirring for 2-3 hours. The solution was then centrifuged at max speed for 10 minutes, the supernatant was collected and brought to pH 8.0 using 10M NaOH. It was then autoclaved and the head space was exchanged with N₂ for ~20 minutes.

First culturing experiments. 20 Hungate tubes were autoclaved and the headspace cleared with N₂ gas aseptically. An equal amount of base was added to each tube (1.8362μL), and the supplements were then added for 2 mL based on (Table 4). The permutations are described in Table 5. A drop of spore stock was added to each sample. Samples were removed with a sterile needle and inspected under the microscope for growth.

Table 5 First culturing attempt supplement permutations

1	Rumen fluid	11	D-glucose, D-xylose, D-mannitol, D-galactose, fructose
2	Algal extract	12	Yeast extract, tryptone, algal extract, D-xylose, D-mannitol, D-galactose, bile salts, urea
3	D-glucose	13	Algal extract, D-xylose, D-mannitol, D-galactose, bile salts
4	D-galactose	14	Algal extract, D-galactose, D-xylose, D-mannitol, bile salts, urea

5	D-mannitol	15	Rumen fluid and D-glucose
6	Fructose	16	Algal extract and D-galactose
7	D-xylose	17	All supplements - urea
8	Yeast extract + trypton	18	All supplements - bile salts
9	Volatile fatty acids	19	All supplements - volatile fatty acids
10	Algal extract, D-glucose, D-xylose, D-mannitol, D-galactose	20	All supplements

Second culture attempt media. Due to the complications with the first media, all media was made in a CO₂ environment. Hungate tubes were no longer used, rather regular sterile culturing tubes were brought into the anaerobic chamber 3 days in advance giving them time to calibrate to the CO₂ environment, and remove all O₂. We used a Coy anaerobic chamber with a 95% CO₂/5% H₂ atmosphere. Media was prepared in a similar fashion except CO₂ was used to exchange head space. The autoclavable solutions were made in 60 mL of water in a Hungate vial, autoclaved, and gas exchanged. The non-autoclaved components were combined together in the remaining water and then sterile filtered into the cooled down vial after gas exchange had occurred in the anaerobic chamber. If the resazurin showed signs of oxygen present, the gas was exchanged longer until the pink color turned yellowish clear. Different combinations of buffers were used; sodium bicarbonate was still used for pH 7.2, while 1M HEPES pH 8.1 and 1M TRIS pH 8.8 stocks were used as buffers at a final concentration of 0.1M. The Good buffers were prepared in sterile, Hungate vials flushed of O₂ with CO₂. The sodium bicarbonate buffer was adjusted to pH 7.2 after bubbling for 20 minutes in CO₂. All stocks and samples were left in the anaerobic chamber. Amino acid stocks were made in 15 mL falcon tubes, and sterilized and filtered into new ones. Hours before media was to be made, they were transferred from storage in the fridge into the anaerobic chamber to calibrate and lose oxygen. The media recipe is in Table 6, with the amino acid stocks in Table 7. To make plates the media was made at a 2X concentration in 50 mLs (the amounts used for 100 mL were added in 50 mL), and 1% agar was

made in 50 mL of water. The two components were quickly combined inside the anaerobic chamber and plates were poured.

Table 6 Second media recipe

Component	Amount to add for 100mL (in grams)	Amount of stock added for 100mL	Stock solutions % (w/v)
NaCl	0.8		
KCl	0.04		
CaCl ₂ 2H ₂ O*	0.02	0.2mL	10
NH ₄ NO ₃ ⁻	0.033		
MgSO ₄ 7H ₂ O*	0.1	1mL	10
KH ₂ PO ₄	0.02		
K ₂ HPO ₄	0.015		
FeSO ₄		100μL	4.00E-02
Na pyruvate	0.02		
Vitamin K3		50μL	1mg/mL in 100% ethanol
vitamin mix		1mL	
trace mineral mix		1mL	
cysteine HCl	0.06		
resazurin		100μL	0.1
Amino acid mix		12.5mL	1mM
Fructose	0.1		
D-mannitol	0.1		
D-glucose	0.1		
D-xylose	0.1		
D-galactose	0.1		
tryptone	0.1		
yeast extract	0.1		
calf thymus DNA		200μL	2.5mg/mL
Algal extract*		2mL	
Buffer	0.1g NaHCO ₃ or	10mL HEPES/TRIS	1M
H₂O		71.85 or 81.85mL	

Bold font indicates those that were not autoclaved. * indicate samples that were autoclaved separately and then added

Table 7 Amino acid stocks

Amino acid	stock concentration	Amount added for	Amino acid	stock concentration	Amount added for 100mL media
------------	---------------------	------------------	------------	---------------------	------------------------------

		100mL media			
A	10 mg/mL	500μL	M	10 mg/mL	500μL
C	4 mg/mL	1mL	N	10 mg/mL	500μL
D	50 mg/mL in 1M NaOH	1mL	P	10 mg/mL	500μL
E	50 mg/mL	1mL	Q	10 mg/mL	500μL
F	10 mg/mL in 0.01M NaOH	500μL	R	10 mg/mL	500μL
G	10 mg/mL	500μL	S	10 mg/mL	500μL
H	10 mg/mL	500μL	T	10 mg/mL	500μL
I	10 mg/mL	500μL	V	10 mg/mL	500μL
K	10 mg/mL	500μL	W	2 mg/mL	1mL
L	10 mg/mL	500μL	Y	2 mg/mL in 0.01M NaOH	1mL

Amino acids are represented by their single letter code. Only the L isomers were added

Growth experiments. Spore stocks from *N. unicornis* fish 2 caught on 4/13/12 were brought into the anaerobic chamber, and 150μL were spun down. The supernatant was removed and the stocks were washed 2x in 600μL sterile dH₂O to remove the residual glycerol. They were washed one more time, and the 600μL was broken up into 2 tubes (300μL in each), spun down, and resuspended in appropriate treatment solution (appropriate buffer for no-treatment, 10mM HCl pH 3 for acid treatment, or 10mM HCl pH 3 with 5mg/mL pepsin for pepsin treatment). The treatments were incubated at 37°C for 30 minutes. Acid and pepsin treatments were spun down, washed with H₂O 2x, and resuspended in appropriate media. 50μL was added to 5mL of media. For volatile fatty acid (VFA) treated samples, 50μL of 50mM VFA solution (contains 50mM of acetate, propionate, isovalerate, and butyrate, pH 7.0) was added. The schematic of culturing attempts are displayed in Figure 1.

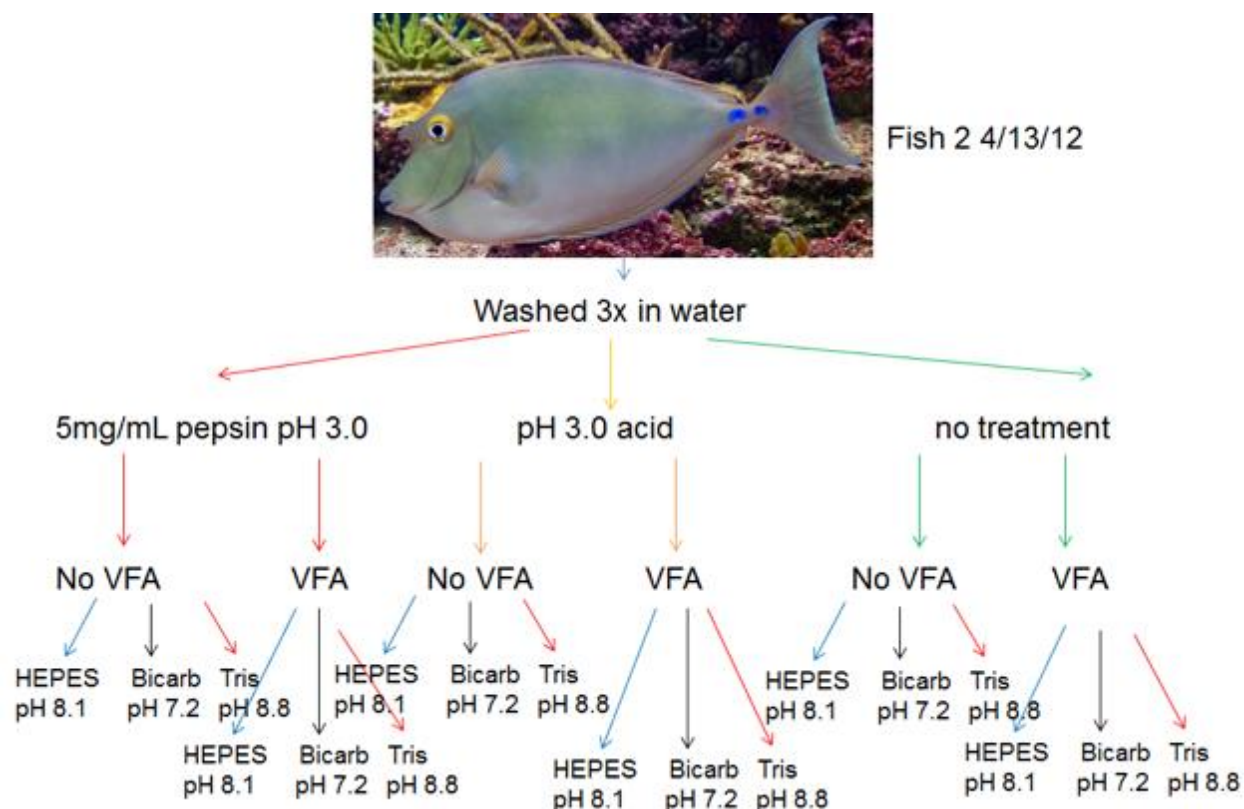


Figure 1. Schematic of experimental treatments. Samples were treated as described in the text. All cultures were done in duplicate with 5mLs of culture.

10 μ L samples were taken out using a p10 pipettor and examined on microscope slides every day for 1-3 weeks. Potentially promising cultures were plated, and colonies were examined microscopically. The 16S rRNA gene sequence was PCR amplified from colony PCR using 8F (AGAGTTTGATCCTGGCTCAG) and 1492R (GGTTACCTTGTTACGACTT) universal primer set. The PCR product was purified and then submitted for sequencing at the Cornell sequencing center. Cultures were potential were subcultured into fresh media after one week post-inoculation by adding 50 μ L to 5mL of fresh media.

Cell staining. Potential epulo containing cultures were DAPI stained following the standard Angert lab protocol, to assess if DNA patterns were reminiscent of what was observed in other epulos. To determine if certain cells were epulos, Fluorescent in Situ Hybridization (FISH) was

conducted using the 1423 *Epulopiscium* spp. specific probe (5), following the standard Angert Lab FISH protocol. The 515 universal probe was used as a positive control. Samples were also tested for viability using the BacLight LIVE/DEAD staining kit (ThermoFisher Scientific) to determine if membranes were compromised and if cells appeared to be dormant or vegetative.

Third set of culturing experiments. The third set was performed on more recent spore stocks from both *N. unicornis* and *N. lituratus*. The only differences were that sodium bicarbonate was the only buffer used, and the carbohydrate concentrations were reduced 10-fold (Table 8).

Table 8 updated media

Component	Amount to add for 100mL (in grams)	Amount of stock added for 100mL	Stock solutions % (w/v)
NaCl	0.8		
KCl	0.04		
CaCl ₂ 2H ₂ O*	0.02	0.2mL	10
NH ₄ NO ₃ ⁻	0.033		
MgSO ₄ 7H ₂ O*	0.1	1mL	10
KH ₂ PO ₄	0.02		
K ₂ HPO ₄	0.015		
FeSO ₄		100μL	4.00E-02
Na pyruvate	0.001		
Vitamin K3		50μL	1mg/mL in 100% ethanol
vitamin mix		1mL	
trace mineral mix		1mL	
cysteine HCl	0.06		
resazurin		100μL	0.1
Amino acid mix		12.5mL	1mM
Fructose	0.01		
D-mannitol	0.01		
D-glucose	0.01		
D-xylose	0.01		
D-galactose	0.01		
Maltose	0.01		
tryptone	0.01		

yeast extract	0.01		
calf thymus DNA		200μL	2.5mg/mL
Algal extract*		2mL	
Buffer	NaHCO₃ 0.1	10mL HEPES, TAPS, or TRIS	1 M
H₂O		81.85mL	

Bold font indicates those that were not autoclaved but rather filter sterilized. * indicate samples that were autoclaved separately and then added

Heat treatment of spores. These spores were also heat-treated in an attempt to activate them for germination. 50μL of spores were washed and resuspended in 300μL of H₂O. They were then heat-treated at 60°C for 45 minutes. Samples were also treated through a Percoll gradient. Spore samples were resuspended in H₂O, then added to the top of 10 mL of Percoll diluted 9:1 in 1.5M NaCl, and spun at 10,000 x g for 15 minutes. The bottom 1 mL was collected and washed 3x with H₂O and used as inoculum after resuspension into the buffer. This was done in hopes to limit the smaller spores from entering the media.

TAPS buffer. It appeared as though the pH of the TRIS buffered media would fluctuate and there were growth complications due to a contaminating bacterium, therefore TAPS buffer pH 8.9 was generated and used for high pH as well. 1M stock was used. Spores were heat treated the same way as previously mentioned and grown in in the TAPS buffered media.

Growth from *Zebrasoma flavescens* feces. *Z. flavescens* feces was collected from the aquarium that the fish were living. The feces was mixed by pipetting up and down, centrifuged and washed in H₂O 3x. It was resuspended in the 4 different buffered media (sodium bicarbonate, HEPES, TRIS, or TAPS). Feces from the bags the fish were delivered in was also used for culturing. The

bags containing the fish were filtered through a 12µm pore Milipore membrane filter, to try to select for large epulo cells. The membranes were added directly to the media tubes for growth.

***A. nigrofuscus* gut extract culture attempts.** *A. nigrofuscus* gut extract was generated after dissecting an *A. nigrofuscus* specimen sent to us from Hawaii. The fully dissected gut excluding the stomach was ground up in a 1.5mL microfuge tube with a sterile pestle. Due to the rubbery nature of the gut, this was difficult and did not completely break apart, but pieces did break off and contents homogenized. This was resuspended in 1mL of water, transferred to a Hungate vial, and 24 mL of dH₂O was added. This was then autoclaved for 15 minutes, and the headspace replaced with CO₂.

B.4 RESULTS

Culturing attempts in N₂ headspace Hungate tubes

The first set of culturing experiments failed for multiple reasons. Many of the tubes failed to stay anoxic and turned pink. Those that remained anaerobic were overrun with a contaminant coccus that dominated the cultures. Though fruitless, the first set of experiments was helpful in streamlining the process of culturing as it showed that performing culturing experiments with multiple permutations was difficult if everything was done in Hungate tubes. The N₂ atmosphere also appeared to have problems staying anaerobic, and had more precipitation problems when using it. From these results it was evident that the culturing needed to be done in the anaerobic chamber to maintain anoxic conditions, and a CO₂ environment is likely better for setting up media and making cultures. Further it showed that to do multiple tests, using regular culturing

tubes already calibrated to a CO₂ environment would likely be more efficient and high-throughput than Hungate tubes.

Culturing attempts in a CO₂/H₂ atmosphere.

Spore stocks were cultured in a way to potentially mimic passage through the gut. There were 18 different possible treatments, as 3 buffers varying in pH were used. The high pH of 8.8 was used because motility of type C cells in buffer at this pH was previously observed. Bicarbonate was used as a neutral buffer (pH 7.2), and HEPES at pH 8.1 was used as a more intermediate buffer. The route of transmission is unclear in epulos in surgeonfish hosts, but experimental evidence of a starved captive *N. lituratus* showed that it had type C endospores in its stomach, suggesting there is a potential oral-fecal route of transmission. Due to this, we mimicked the potential stomach passing by treating spores either in an acid treatment (pH 3.0) or an acid and pepsin treatment. We also used no treatment controls, to see if there was any effects of the low pH or pepsin. These were done for media buffered with each buffer. Each of these were further broken down into treatment with or without VFAs to potentially stimulate growth. We observed differences in the cultures based on buffer used, but spore treatments and the presence or absence of VFAs did not appear to have any effect on culture composition, or provide any advantage to growing epulos. In each media, there were no motile epulos, and no real evidence of growth, but there were some potentially viable cells. This was assessed through the presence of phase-dark cells in the media, as well as through LIVE/DEAD staining, which assesses membrane integrity. This is discussed below for each media type.

HEPES buffered media

All the HEPES buffered media supported a large abundance of growth, regardless of treatment. Highly motile cells were found in each culture, and these included species like

Clostridium oceanicum, which produced 2 endospores at each pole, and other endospore formers with the clostridial cellular form (Figure 2). All HEPES cultures produced phase-dark type C cells, as well as phase-dark type J cells. From inspection of the spore stocks, phase-dark cells of both morphotypes were rarely present, suggesting that these cells germinated from spores in the media. The phase-dark type Cs had a rough appearance to their surfaces, this suggested that these may have been viable. These types of cells were viewed 2 weeks post-inoculation, but there did not appear to be an increase in abundance of any of the epulo morphotypes. If germination was occurring, it appeared as if the cells germinated, but did not proliferate. Some of the type C cells showed intricate patterning at the cell surface, but it is unclear what it is, or why it occurred. Subculturing the cultures did not promote the growth of any of the epulos either, but some phase-dark and phase-bright epulo cells were present in the subcultures. No binary fission was observed of any potential epulo in these first cultures. Plating did not help culture any epulos either, and plating was made difficult as some plates were eventually destroyed because of colonies producing agarases. Representative cells are present in Figure 2. Results from the HEPES media suggests that it supported the germination at low levels of certain epulos.

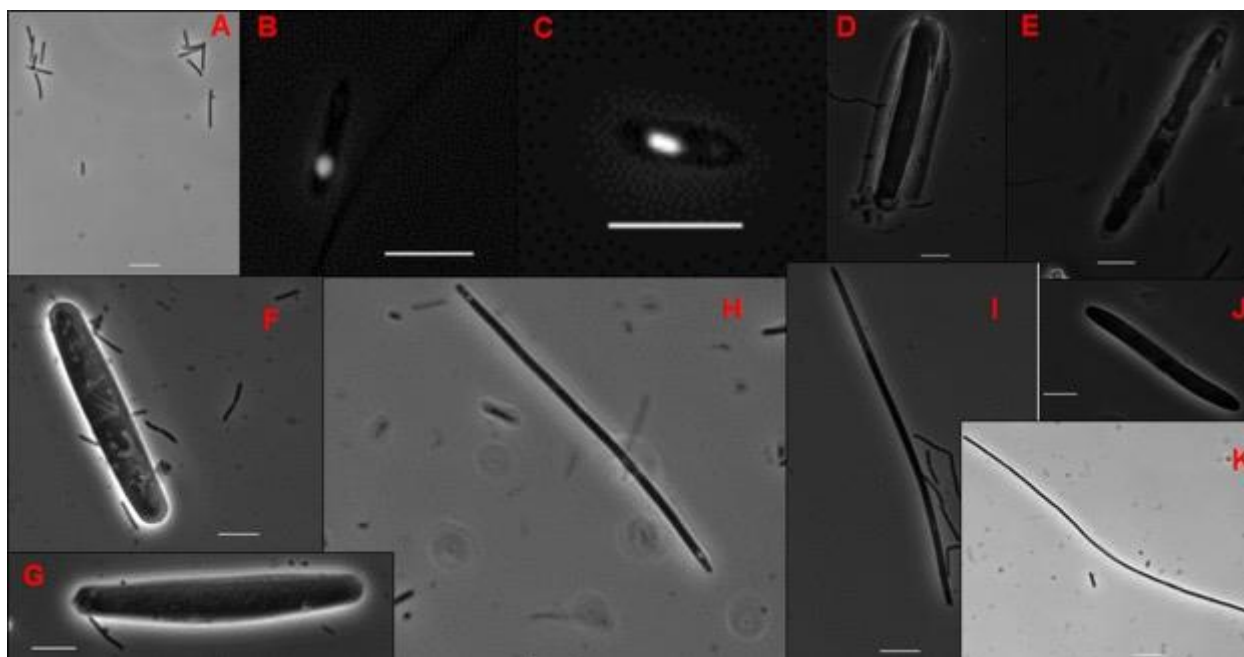


Figure 2 Representative cells from the HEPES culture tubes. Scale bars in each image represent 10 μm . (A), (D), (F), (G), and (K) all show pepsin treated cells. (B), (C), and (E) were acid treated cells. (H) was treated with pepsin and had VFAs in the media. (I) was not treated, but had VFAs present, and (J) had no treatment. (A) shows representative rods found in the culture at 5 days, these were motile. (B) shows a *Clostridium oceanicum* cell at 7 days post-inoculation (C) shows a potential clostridial cell that is lemon shaped 7 days post-inoculation. (D) depicts a phase-dark type C spore and (E) shows 2 phase-dark C2 spores that potentially germinated in the media at 7 days post-inoculation. (F) shows a phase-dark type C cell 7 days post-inoculation with complex patterning on the surface. (G) shows a phase-dark type C cell with a rough exterior that likely germinated 12 days post-inoculation. (H) and (I) show phase-dark type J cells with potential endospores forming at their poles. (H) was viewed 13 days in the media, while (I) was only in the media for 2 days. (J) shows a potentially germinated phase-dark epulo. (K) shows a long phase-dark type J present in the subcultured tube after 6 days post-inoculation.

TRIS buffered media

The TRIS buffered media appeared to be the worst for culturing epulos, as the cultures were very quickly overrun by a certain bacteria. This caused the cultures to take on a large clumpy appearance. This was caused by a bacterium, likely *Oceanirhabdus sediminicola* that formed these rope-like clusters (Figure 3). This made investigating the cultures difficult as these clumps were large and the cells were motile. Despite this, some potential germination of epulos was witnessed in the cultures. Plating did not provide any help either, and the rope-like bacterium did form isolated colonies on the plates. Due to the abundance of and complications caused by this

bacterium, the TRIS buffer was the weakest buffer to use for attempting to culture epulos from spore stocks.

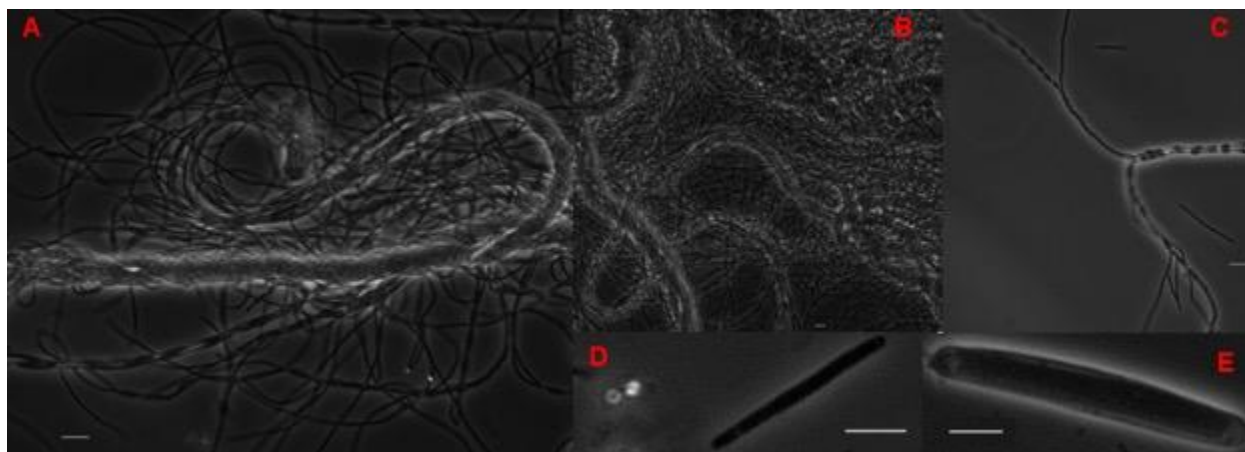


Figure 3 Representative cells from the TRIS culture tubes. Scale bars in each image represent 10 μm . (A-C) were all 2 days post-inoculation. (D-E) were 6 days post-inoculation. (A) shows a cluster of the rope-like bacteria from the pepsin treated spore stock. (B) shows a larger view of another cluster of rope-like bacteria in the culture from the pepsin treated spore stocks with VFAs. (C) shows the formation of a clump of rope-like bacteria in the non-treated culture, 7 days post-inoculation. (D) shows a potential phase-dark epulo that might have had a spore forming at the top pole 7 days post-inoculation. (E) shows a phase-dark type C that likely germinated in the media and may have had endospores forming at the poles. This was found in the VFA treated media.

Bicarbonate buffered media

The bicarbonate buffered media appeared to have worked the best for attempting to culture epulos, although none were definitively cultured. All culture tubes supported a large abundance of growth of other spore formers. There were no differences in treatment on the spores, but in many cases, it appeared as if germination occurred. There were also multiple phase-dark long skinny cells that could have potentially been epulos present in the cultures and they seemed to be more abundant in this media than any of the other buffered media. Due to this, this culture was investigated through LIVE/DEAD staining to see if any of these potential epulos were viable. Some of the phase-dark cells did have staining patterns consistent with cells that had intact membranes (Figure 4E), suggesting that they were potentially viable in the media, though

they were not motile. Subculturing did not improve any epulo germination or growth attempts, though phase-dark epulos were present in subcultures. The media also produced these long, skinny cells that resembled type J cells. These cells were motile and appeared to rotate in a sinusoidal fashion. Pure cultures of these cells were grown after the culture was plated and colonies were isolated. The cells produced these wave-like growth patterns on the plates that appeared to move. The cells were DAPI stained to see if their staining patterns were consistent with other epulos (Figure 5A), as well as probed via FISH with bacterial and epulo specific probes. The DAPI patterning was quite uniform throughout the cells, with some spots of higher fluorescence intensity. The cells did not stain when provided the epulo probe, suggesting they were not epulos. This was confirmed as sequencing of PCR products of their 16S rRNA genes came back as *Oceanirhabdus sediminicola* hits after BLAST analysis. Due to bicarbonate appearing to be the best media, it was used and modified in further experiments.

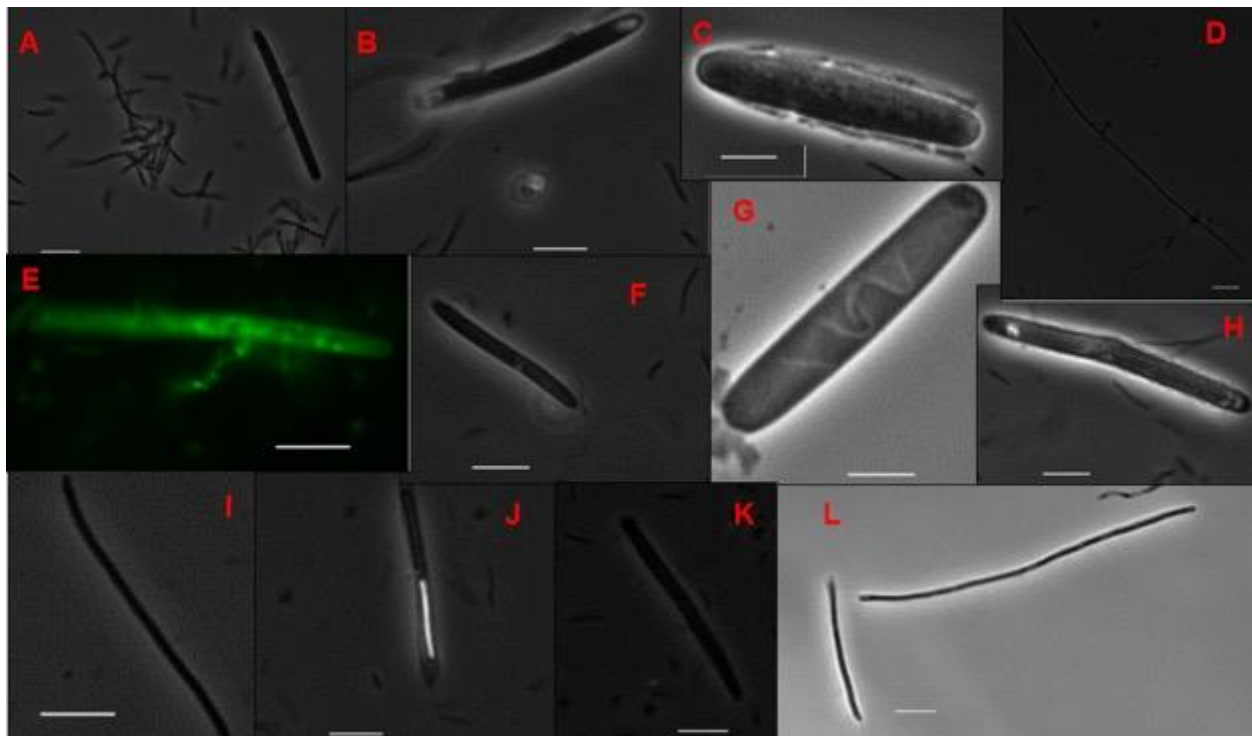


Figure 4 Representative cells from the bicarbonate culture tubes. Scale bars in each image represent 10 μm . (A), (C), and (E) were pepsin treated spores supplemented with VFAs, (B) and (D) were untreated cells. (F), (G), and (J) were pepsin treated, (H) and (K) were acid treated

supplemented with VFAs, while (I) was acid treated. (L) was the pure culture of the potential epulo. (A) shows the type of cells present at 7 days, with a phase-dark epulo like cell present. (B) shows an epulo at 13 days post-inoculation that is phase-dark with spores developing at the poles. This cell appears to be viable. (C) shows a phase-dark type C cell potentially with either the dead mother cell or spore cortex around the cell. The image was taken at 5 days, and the surface appears rough and textured. (D) shows a type J that is phase-dark after 7 days post-inoculation. (E) depicts a LIVE/DEAD stained potential epulo at 5 days. The DNA was stained and the cell was phase-dark, suggesting that the cell was viable. The propidium iodide did not penetrate the membrane demonstrating it was not compromised. (F) and (K) show similar types of potential epulos that are phase-dark. (F) was imaged at 12 days, while (K) was at 7 days. (G) shows a type C cell that appears to have germinated at 2 days post-inoculation. There was a helical pattern in the cell, interestingly, the DNA nucleoid is helical in type C's suggesting this could possibly be DNA or some sort of scaffolding for the DNA nucleoid. (H) shows 2 phase-dark spores that were present in a mother cell at 1 week. (I) shows a potential type J at 12 days. (J) shows an epulo with 1 phase-dark and 1 phase-bright spore, suggesting one germinated. (L) shows the potential epulo in pure culture, later determined to be *O. sediminicola*.

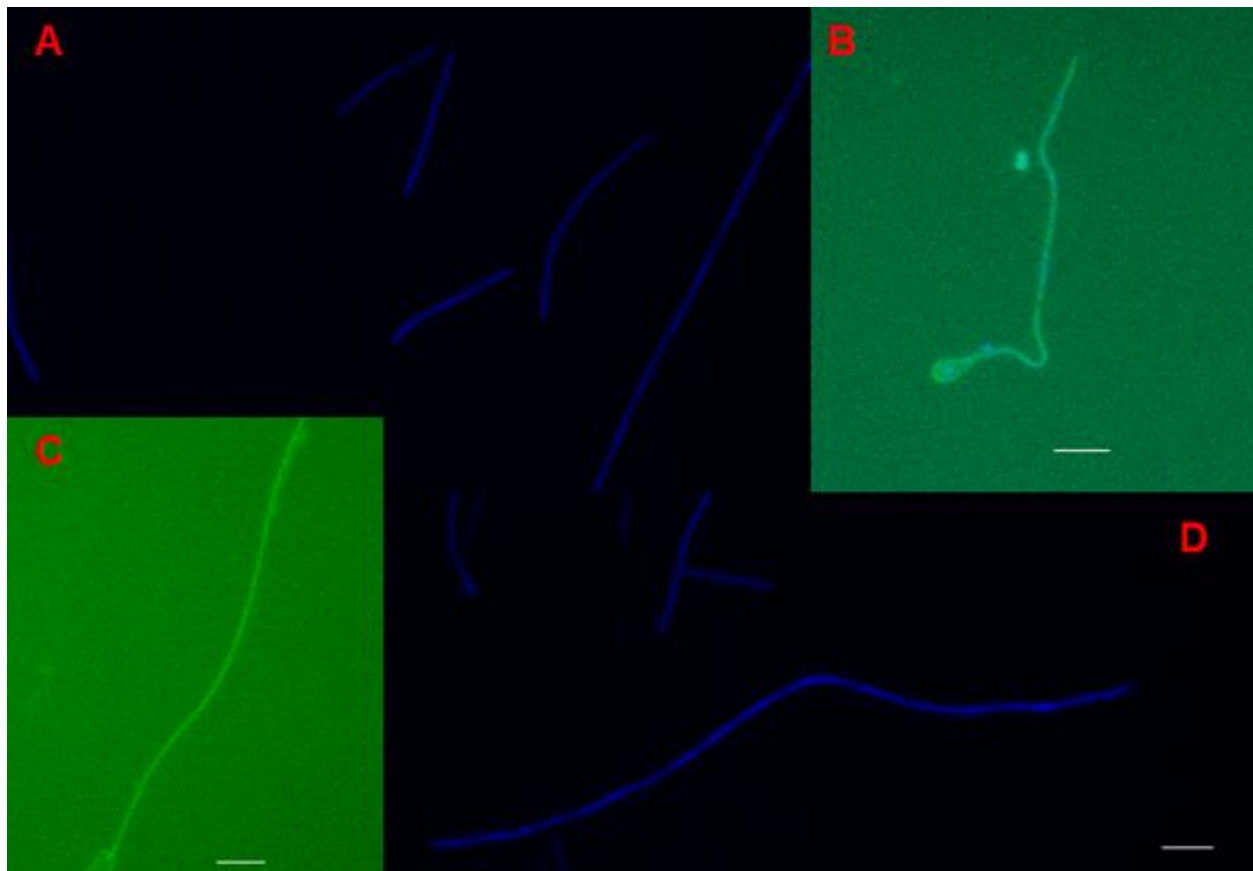


Figure 5 DAPI and FISH staining of the potential epulo found in bicarbonate media. All scale bars are 10 μ m. (A) and (D) show generally uniform DNA staining from DAPI. (B) shows the potential epulo stained with the epulo specific FISH probe. There is not much fluorescence present, suggesting the probe did not hybridize. (C) shows the potential epulo hybridized with a universal bacterial FISH probe. There was fluorescence, suggesting that this bacteria was not an epulo.

Culturing from new spore stocks

Due to the lack of conclusive evidence of growth from the old spore stocks, but promise with the media due to germinating cells, we tried culturing from new spore stocks collected from more recently caught *N. unicornis* and *N. lituratus* guts. This was in hopes that the more recent stocks would be more amenable to culturing. Further we updated the media recipe by reducing the carbohydrate concentration 10-fold in an attempt not to overwhelm the epulo cells. These cultures were unsuccessful in growing epulos, but they did show more of these dim phase-bright cells with textured surfaces that were not abundant in the previous experiments. Again there was no difference in acid, pepsin, or no treatment, or the presence or absence of VFAs. Only bicarbonate was used at this period due to it appearing to be more successful than TRIS or HEPES. There was one case of a potential epulo undergoing binary fission in the VFA supplemented *N. lituratus* treated sample at 8 days post-inoculation (Figure 6K). This was the only time this was observed and there was no increase in abundance of the cell type, making it impossible to confirm if this cell was an epulo or not. LIVE/DEAD staining also revealed that some of the potential epulos appeared to be viable. The cultures are summarized by representative images in Figure 6.

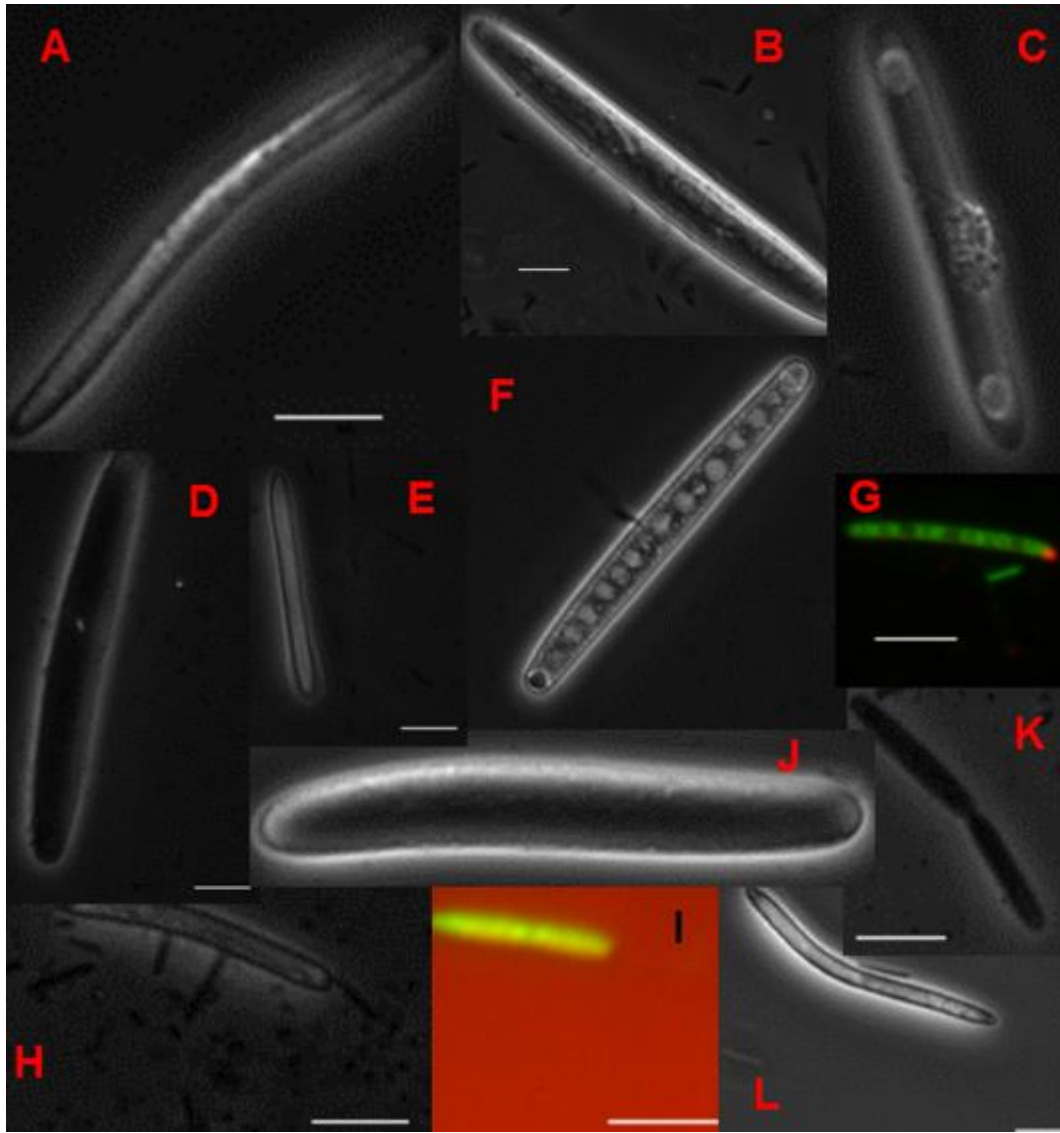


Figure 6 Representative cells from the bicarbonate culture tubes inoculated with *N. lituratus* spore stocks. Scale bars in each image represent 10 μm . (A) was treated with acid, (B) and (F) were treated with pepsin and supplemented with VFAs, (C), (G), and (J) were untreated, and (D), (E) and (I) were treated with pepsin. (H) was acid treated and supplemented with VFAs, and (K) and (L) were untreated but supplemented with VFAs. (A), (E), (L), and (H) illustrate the dim phase-bright spores present in the media. It is unclear the status of these cells or if they were viable. (A) was taken 16 days post-inoculation. (B) was imaged at 2 days and shows a phase-dark type C that appeared textured, suggesting it germinated. (C) and (D) show similar cells that were both phase-dark and at different stages of development. (C) was taken at 8 days and had endospores developing, while (D) was taken at 2 days. This suggests that the cells might have been undergoing sporulation. (E) was taken at 16 days. (F) was taken at 18 days and shows a type C with this intricate patterning. It is unclear if those are blebs due to lysis. (G) shows a potentially viable epulo with patchy staining from the LIVE/DEAD staining. This suggests the cell may have been viable. (H) was taken at 8 days, while (I) shows a potentially viable cell based on LIVE/DEAD staining at 2 days. (J) shows a potentially viable, phase-dark, textured type C cell at 21 days post-inoculation, suggesting the cell likely germinated and was able to

survive in the media for up to 3 weeks. (K) shows a potential epulo undergoing binary fission at 8 days, and (L) shows a potential epulo at 2 days post-inoculation.

The new *N. unicornis* samples also had an increased abundance of the dim phase-bright epulo cells. These cells tended to stain live in LIVE/DEAD stains in a uniform pattern. It is unclear if this is because they were still endospores and the SYTO9 was attracted to their coats, or if they were viable cells. The intensity and staining patterns of more clearly discernable endospores differed, suggesting that these dim phase-bright cells were vegetative. The untreated *N. unicornis* culture appeared to potentially be promoting the growth of an epulo. These potential epulos appeared to germinate and there may have been an increase in abundance of these cells early in the culture, however, as the culture progressed, the cells were no longer prevalent. It is unclear what happened to them, but one cell was found undergoing binary fission (Figure 7F). Plating attempts did not allow for the isolation of this potential epulo.

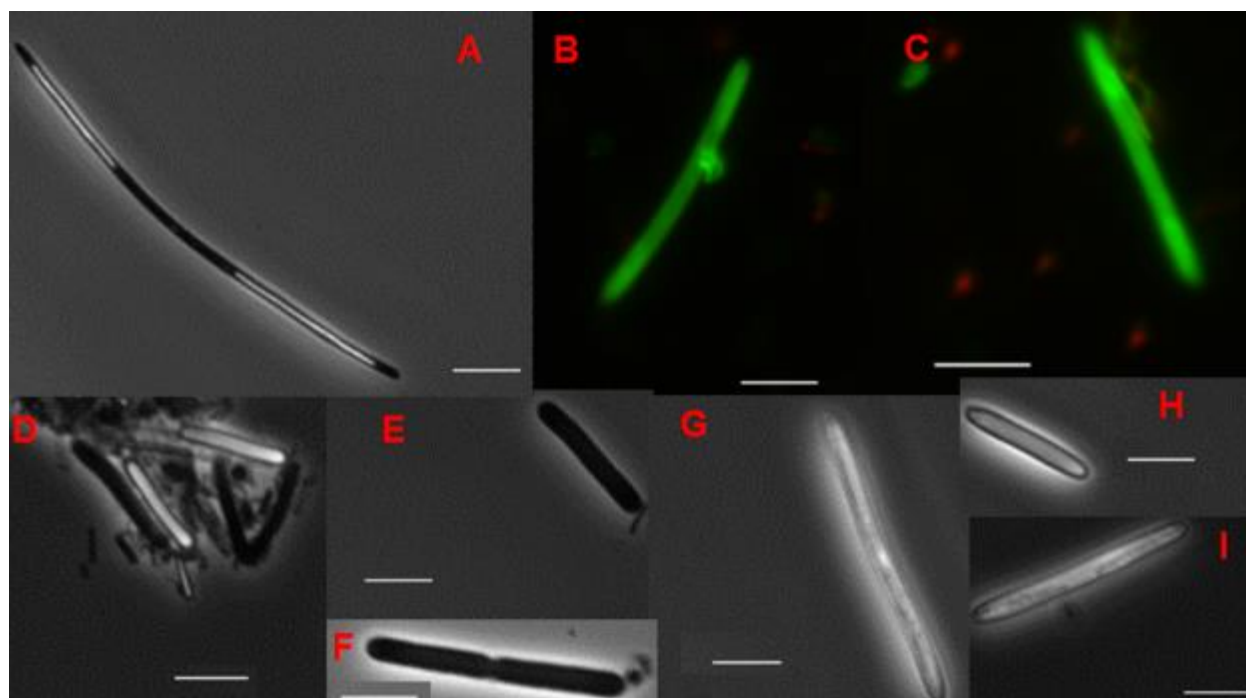


Figure 7 Representative cells from the bicarbonate culture tubes inoculated with *N. unicornis*. Scale bars in each image represent 10 μm. (A) and (G) were from pepsin treated spores, (B) and (C) were from acid treated spores, (D-F) were untreated but supplemented with VFAs, and (H) was pepsin treated and supplemented with VFAs, while (I) was not treated. (A) shows a type J cell at 2 days with endospores present at the poles. The mother cell was phase-dark suggesting it

may have been viable. (B) and (C) show 2 epulos stained through LIVE/DEAD staining being potentially viable based on staining patterns. More SYTO9 staining occurred where the daughter cells were forming. (D-F) show the same type of potential epulo. (D) was taken at 2 days and shows a group of epulos with some phase-dark cells appearing to have germinated while other cells are still phase-bright endospores. (E) shows a single cell of the potential epulo, while (F) shows one undergoing binary fission, both (E) and (F) were taken at 1 day post-inoculation. (G-I) depict the dim phase-bright cells. (G) was taken at 21 days, while (H) was at 2 days and (I) was at 1 day.

Heat treatment

Due to the lack of any clear growth of epulos, we tried a new activation method, in which spores were heat treated at 60°C to potentially help spawn germination and growth. This was performed on all 3 spore stocks. This did not improve culturing from the other treatments, but in each spore stock, there was a continued abundance of the dim phase-bright cells. These were stained with LIVE/DEAD stain and showed high fluorescence with the SYTO9 dye. There appears to have been germination in this media, but there was no apparent growth. We also tried selecting for epulos by running spores through a percoll gradient to pellet the epulo spores and remove some of the other smaller contaminating spores. This did not improve anything and did not really select for larger epulos, as smaller cells still were able to germinate and grow in the media.

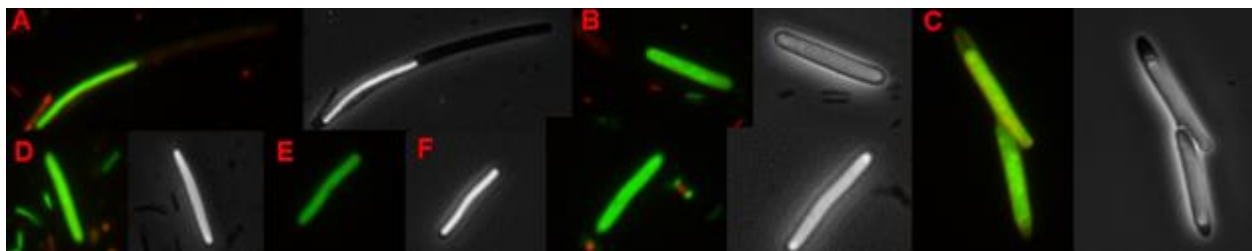


Figure 8 Representative cells from the LIVE/DEAD staining of heat treated spores. (A) shows the staining and bright field image of a type J cell. The phase-dark mother cell did not stain well, however, its endospore had a bright uniform staining. This suggests that the mother cell lost its DNA content. (B) represents the staining pattern of the dim phase-bright cells frequently appearing in these cultures. The bright field image appears textured and this is reflected with a textured DNA staining pattern. There are certain spots with higher staining potentially mimicking the surface of the cell. This suggests that the stain is either attracted to the cell surface or possibly the presence of the nucleoid at the periphery of these cells contributes to their textured and dim phase-bright appearance. A similar staining pattern was observed in (C) which contained cells emerging from a mother cell. (D) and (F) were smaller potential epulos with a

dim phase-bright nature. Their staining patterns mimicked what was observed in the others, as they appeared textured like their cell surfaces. (F) shows a phase-bright endospore with a consistent staining throughout the whole spore. This staining pattern was present in some other phase-bright endospores, distinguishing them from the staining patterns of the dim phase-bright cells.

TAPS buffered media

Due to the complications of working with the TRIS buffered media, we tried another Good buffer for generating high pH media. For these sets of experiments we utilized TAPS at a pH of 8.9 and attempted to grow heat-treated cells in this media. It seemingly allowed for the germination of endospores but there were no clear indications of growth of any epulos or any real improvement of the cultures. This media supported a large abundance of growth of other cells. The TAPS media did not allow for the growth of the rope-like bacteria plaguing the TRIS media. This media did produce one instance of an endospore forming epulo undergoing binary fission (Figure 9I), however, numbers of these cells did not increase in the cultures.

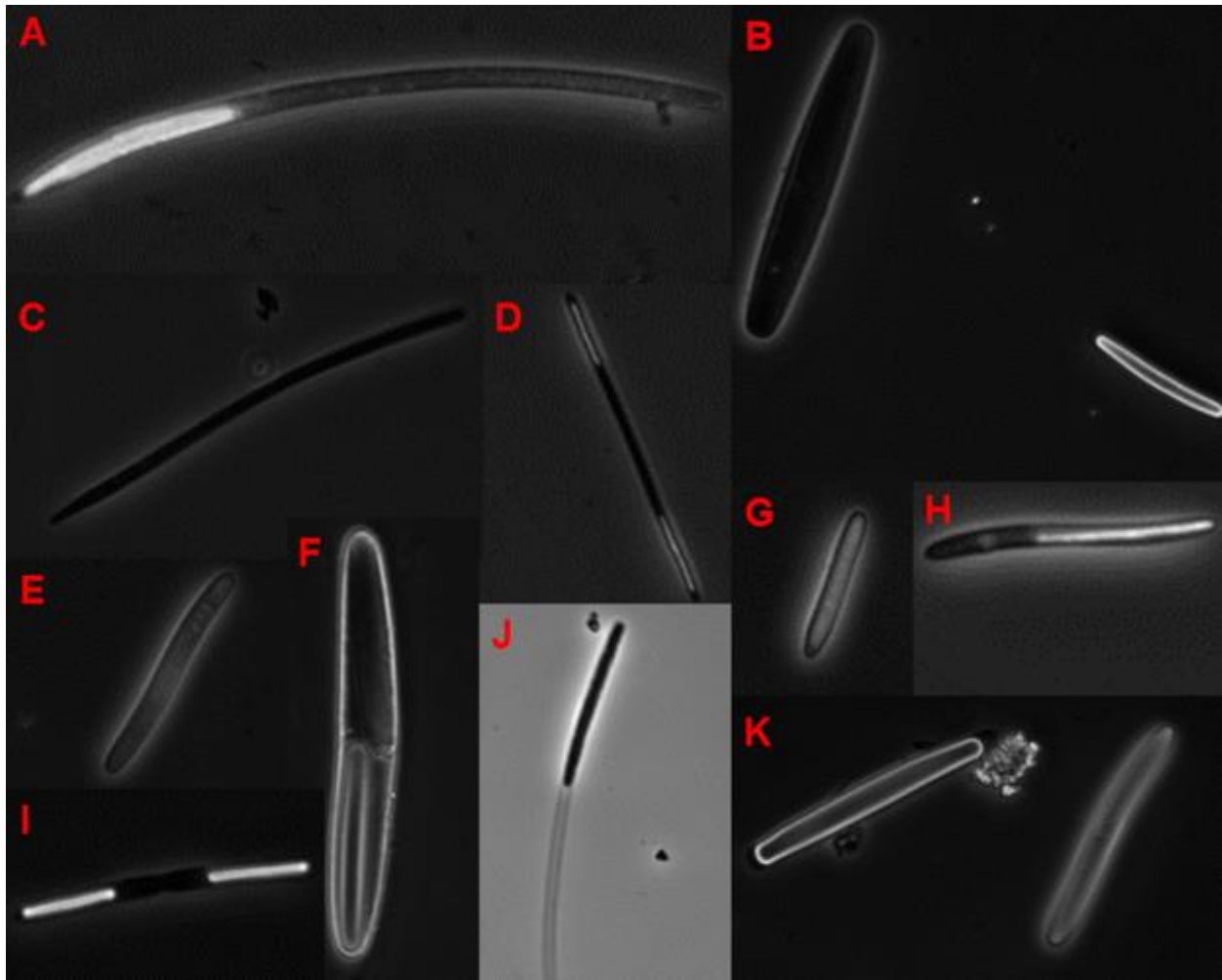


Figure 9 Heat treated spore stocks growing in TAPS buffered media. This figure shows the diversity of epulo cells found in the culture and their different states. (A) shows a type J cell with an endospore at one pole. The mother cell was phase-dark making it appear as if the mother cell was still viable. (B) shows a large type C cell that appeared to have germinated in the media. (C) and (D) depict type J cells, with (C) being completely phase-dark and (D) being phase-dark with developing offspring. (E) demonstrates a dimmer epulo cell. (F) shows a phase-dark cell exiting a type C cell. (G) shows a dim phase-bright potential epulo cell, while (H) demonstrates a small epulo that had an endospore present. (I) shows a small epulo with an endospore undergoing binary fission. (J) shows a degraded mother cell with a phase-dark spore present, suggesting the spore germinated in the media. (K) shows a large type C spore losing its phase-bright appearance, suggesting it might have been germinating. The other cell present in (K) shows a dim phase-dark cell potentially with developing spores at the poles, suggesting it may have been sporulating in the media.

Laser-capturing culture attempts

Type C and type J spores were laser-captured to eliminate the overgrowth of contaminants within the media. 50 spores of each type were collected in 1.5mL microfuge tubes.

The spores were then mixed with the media in Table 8 in the anaerobic chamber, and used to inoculate 5 mL of media. Since there was such a small starting inoculum, investigations into growth were generally fruitless as no cells were taken up on the slide. Eventually the cultures were left alone to see if any turbidity occurred, but this was not the case. No cells were viewed microscopically when investigating the culture. This suggests that the starting inoculum was either too low to see any potential growth, the spores were stuck to the Eppendorf tube and did not transfer to the media, or that the media did not support the growth of these cells directly from spores. No other cells types grew suggesting this method at the very least was effective at preventing the growth of other organisms.

***Zebrasoma flavescens* culturing attempt**

The *Z. flavescens* fecal samples did not contain many epulo-like cells when investigated prior to culturing. Culturing attempts were unsuccessful for attempting to grow epulos, as every culture was quickly overran within 24hrs by other motile bacteria. There was no difference between culturing directly from feces or from the 12µm membrane, suggesting that the size exclusion (selecting for larger cells) did not exclude smaller cells and select for potential epulos.

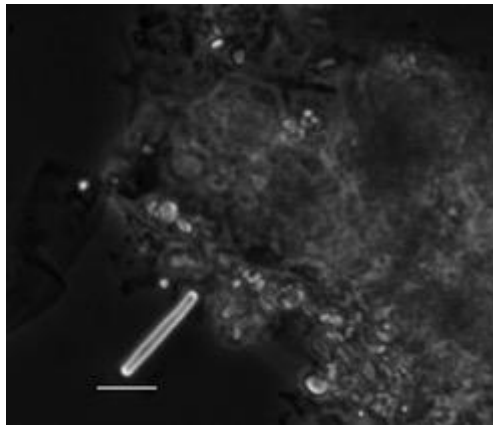


Figure 10 Epulo spore from a fecal sample of *Z. flavescens*. The scale bar is 10 µm. The cell appeared to be an endospore due to its phase-bright nature. These cell types were not abundant in the fecal samples, however, the presence of this cell in the fecal sample suggest *Z. flavescens* contained epulos and sheds them in its feces.

Culturing attempts using *A. nigrofuscus* ‘gut extract’

Due to the lack of any definitive growth and the complicated life cycles of epulos, we reasoned that there may be intrinsic host factors needed to culture cells. Due to the lack of availability of surgeonfish, and the lack of fluidity in their guts, it was difficult to acquire gut fluid and gut extracts analogous to rumen fluid supplements. However, we were able to acquire a few small surgeonfish. These surgeonfish had their guts extracted and observed for the presence of epulos. Although no epulos were visualized in the gut contents, we still utilized their guts to make an extract, in case there were some intrinsic host factors present that would enhance germination or growth. This extract was applied to bicarbonate, HEPES, TAPS, and TRIS buffered culture media. Cultures were investigated microscopically and FISH was performed on them as a test for viability, as it demonstrates the presence of ribosomes within the cell. Still this did not produce any discernable increase in epulo cell mass or any motile cells but there was an abundance in growth in other spore stock members. This was the last culturing attempt done from spore stocks. There was no real success in culturing from these methods, suggesting culturing in the field from vegetative cells would be more successful.

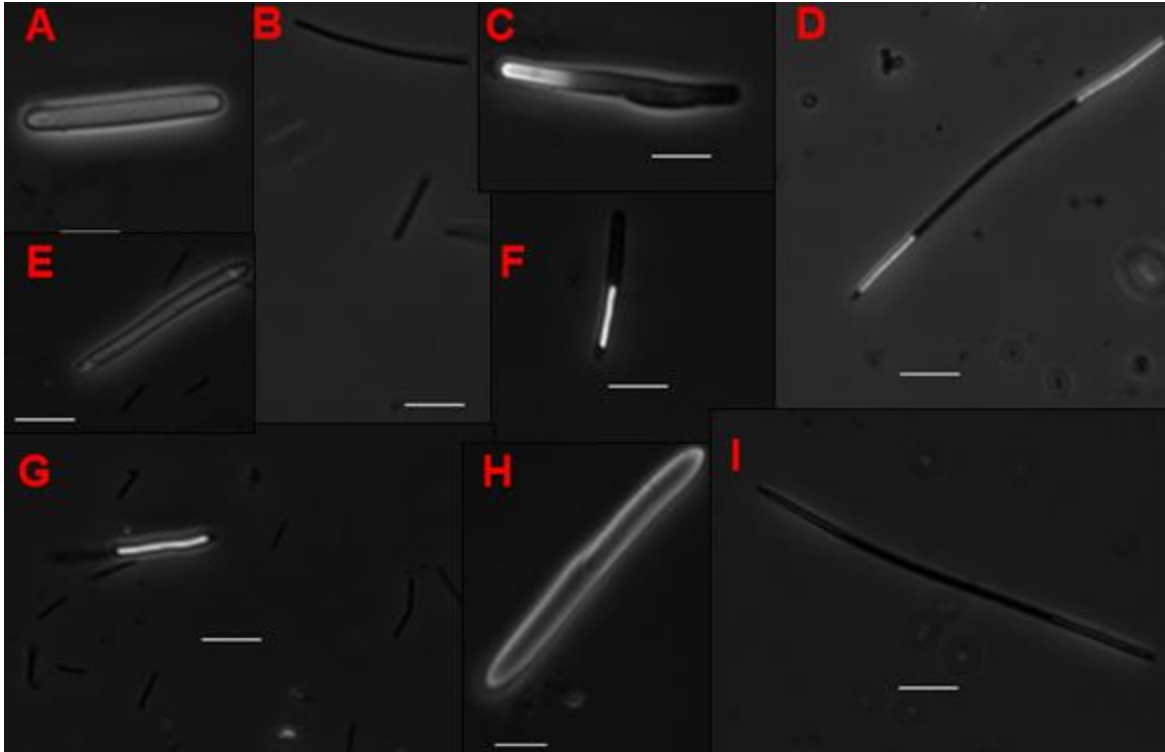


Figure 11 Cells found in cultures supplemented with *A. nigrofuscus* gut extract. (A) cells in bicarbonate media, (B) and (G) are from TRIS buffered media, (C), (E), and (F) are from TAPS buffered media, and (D), (H), and (I) are HEPES buffered media. (A) and (H) show these dim phase-dark spores. (B) and (G) show representative cells that were growing in the cultures. (C) shows a spore potentially leaving dormancy as part of it is phase-bright while the other is phase-dark in the TAPS media. (D) and (I) show 2 type J cells at different stages. (D) shows a cell that had phase-bright stores, while (I) had spores that were turning phase-dark and potentially germinating. (F) shows a small epulo with a phase-dark and a phase-bright offspring cell, suggesting one germinated in the mother cell.

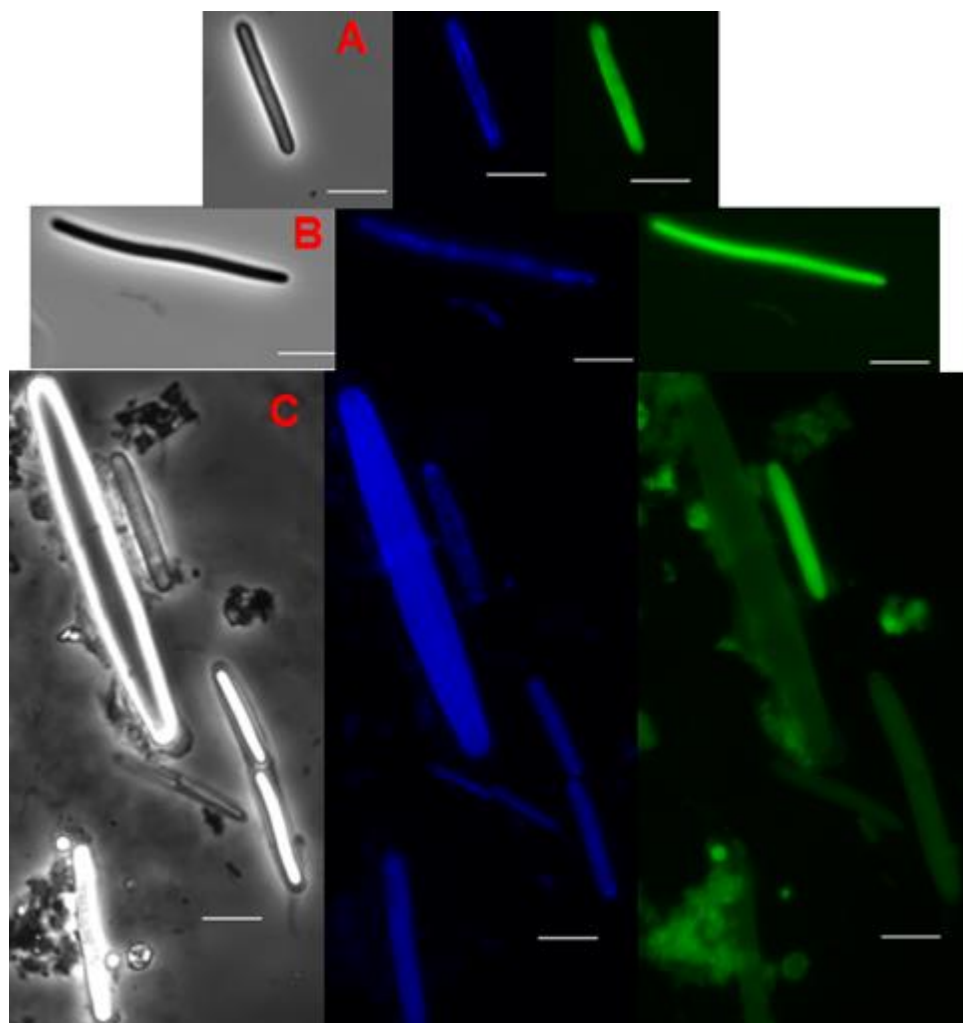


Figure 12 DAPI and FISH staining of epulos in the *A. nigrofuscus* 'gut extract' supplemented cultures. Scale bars represent 10 μm , and images show bright field, DAPI, and FISH stained cells with epulo specific probes. (A) shows a dim phase-bright cell, its DAPI staining pattern and FISH staining pattern. The DAPI staining showed staining at the peripheries with some patterning throughout, and the FISH staining shows pretty consistent staining throughout the cell. This suggests that the cell was viable in the media as its DNA staining pattern was consistent with vegetative cells, and had consistent FISH fluorescence. (B) Shows a phase-dark potential epulo with splotchy DNA staining throughout the cell. The FISH fluorescence was consistent throughout the cell, suggesting the cell was viable in the culture. (C) shows a few types of epulos within the TRIS culture. Included are a large type C spore, a small, dim phase-bright cell next to it, then other smaller epulos. The large type C had consistent DAPI staining indicative of spore staining, but the cell has no FISH staining. The other phase-bright cells had consistent DAPI staining, but no FISH fluorescence, confirming their endospore nature. The small C with phase-bright spores did appear to have extra DAPI staining at the poles. The dim phase-bright cell showed staining at the periphery with extra staining at the poles. FISH fluorescence was consistent throughout the whole cell, suggesting this cell and these types of cells were viable within the culture.

B.5 Section 2: First Culturing attempts at Lizard Island

Media was prepared for Verena Salman to take to Lizard Island to attempt to culture live *Epulopiscium* spp. The media contained the base of Table 8. This media was not supplemented with any VFAs. Media was generated in a Hungate vial and transferred to sterile, sealed, Hungate tubes with a CO₂ environment. There were 4 types of tubes for all 4 buffers, one with the original media (Table 6), one with the reduced sugar (Table 8), and one supplemented with 100 µL of rumen fluid, and one supplemented with 100µL of *A. nigrofuscus* ‘gut extract.’

Some tubes lost their ability to stay sterile and anaerobic in transit and these were discarded and not used for inoculation. Gut samples from *N. unicornis* were used as inoculum for all the tubes. The tubes were investigated microscopically for growth, and were also investigated for overpressure due to the production of volatile fatty acids.

B.6 RESULTS AND CONCLUSIONS

The media that made it to Lizard Island without contamination or losing its anoxic environment were inoculated with gut contents from *N. unicornis* M610. Six hours after inoculation, the *Epulopiscium* spp. were no longer motile in the cultures, however, there was overpressure in some of the cultures, indicating that the cultures were metabolically active. One day post-inoculation, the culture tubes buffered with 100mM HEPES, pH 8.1 were overrun with non *Epulopiscium*-like cells and had overpressure. These cells tended to be non-motile, short rods (Figure 3A) and the type C cells tended to appear non-viable as they lost their cytoplasmic contents, giving them a ghostly appearance. However, in some of these cultures, despite being non-motile, the large type C cells appeared to be normal, but these tubes lacked overpressure, indicating that the cultures probably were not highly metabolically active. The HEPES tube

supplemented with autoclaved *A. nigrofuscus* gut content appeared to have type C cells progressing to the early stages of sporulation (Figure 3B). A similar pattern was observed in the tubes buffered with 100mM TAPS pH 9.0, however, none of these cultures appeared to have cells progressing to the early stages of sporulation. The cultures buffered with 0.1% sodium bicarbonate were generally contaminated and the *Epulopiscium* spp. cells did not appear viable as they had a ghost-like appearance. The sodium bicarbonate culture tube with reduced sugar concentrations had overpressure, with little growth of non-epulo cells, and the large type C's appeared to have begun the process of sporulation (Figure 3C). The culture tubes buffered with 100mM TRIS at pH 8.8 had the healthiest looking *Epulopiscium* spp. cells. Those supplemented with the original, low sugar, and *A. nigrofuscus* had healthy looking large type C cells with little to no cells with a ghostly appearance and no contamination. The type C cells were dark with a reflective white cell periphery, and these tubes lacked overpressure.

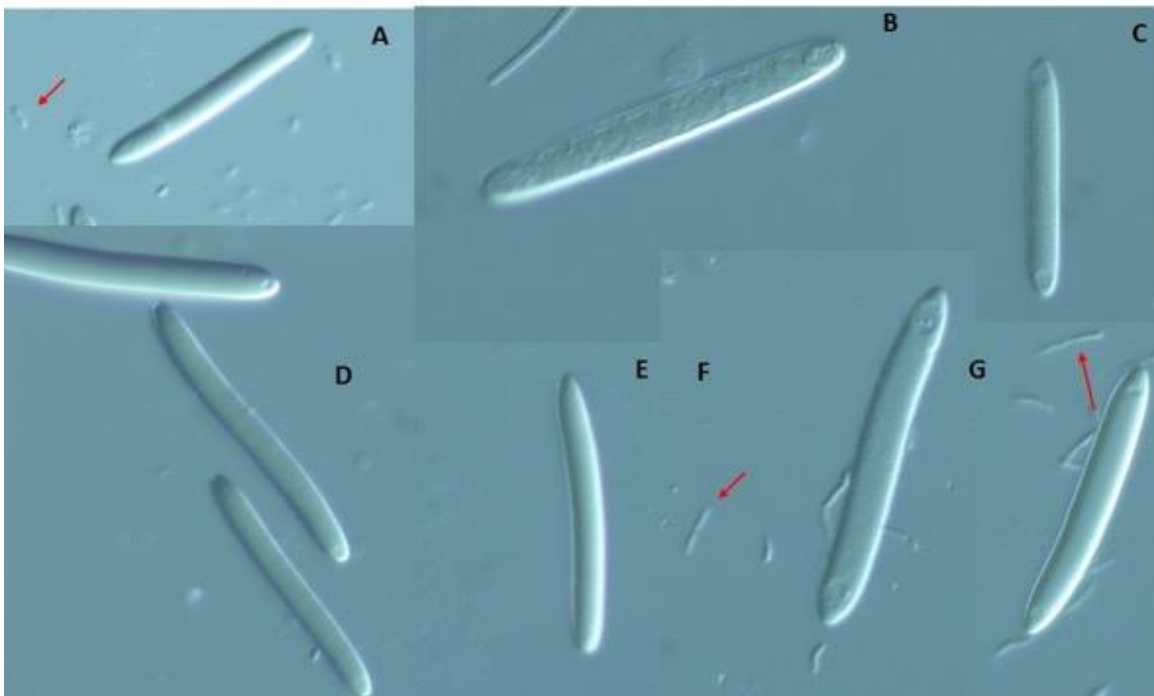


Figure 13. Micrographs of various cultures inoculated with vegetative *Epulopiscium* spp. from *N. unicornis*. (A) shows a type C cell surrounded by smaller, non-motile rods often occurring in pairs (red arrow) in the TRIS supplemented with *A. nigrofuscus* gut contents culture tube. (B)

shows a type C cell in HEPES buffered media containing *A. nigrofuscus* gut contents, with a rougher exterior with forespores present at the poles. (C) depicts a healthier looking type C cell in the reduced sugar sodium bicarbonate medium with forespores present at the poles. (D) and (E) show healthy looking type C cells in TRIS buffered media with original sugar concentrations (Table 8 recipe) respectively. The cells were dark with a very reflective periphery. (F) and (G) show type C cells in the reduced sugar, TRIS buffered media. The cells were darker, with a reflective periphery, and are in the process of sporulating, with forespores at their poles. Arrows are pointing to the motile, skinny rods also present in the culture.

48 hours post-inoculation, the *Epulopiscium* cells tended to no longer appear viable in most of the media. The epulos in the HEPES, TAPS, and bicarbonate buffered cultures that previously appeared healthy and to be entering sporulation were now dead, as they took on a ghost-like appearance. However TRIS buffered cultures were more favorable for the *Epulopiscium* cells. The *Epulopiscium* spp. cells in the original sugar and newest media tubes still appeared healthy as there were no ghostly cells present. These tubes lacked over pressure and lacked the growth of other community members. The reduced sugar TRIS culture tube had no overpressure, but a high density of healthy looking type C cells, however, there was the emergence of long, motile, slender rods in the culture tube, discriminating it from the previously seen short, non-motile contaminants. It is possible that these cells forms a syntrophic relationship with *Epulopiscium* spp. type C, allowing for the higher density of cells in the tube.

The results were inconclusive if any of the cells were actually growing within the media as unfortunately, the *Epulopiscium* cells were not motile. These results suggest that of the buffers tested 100mM TRIS at pH 8.8 was the best buffer for maintaining *Epulopiscium* spp. cells and that type C cells may be viable in co-culture with another community member. The results further show that the bicarbonate buffer the type C cells were usually ghostly in appearance informing us to not use it for attempting to culture type Cs. The results also seemed to suggest

that there was no difference in the growth when different concentrations of carbohydrates were added.

B.7 Section 3: Culturing attempts from multiple surgeonfish at Lizard Island.

B.8 MATERIALS AND METHODS

Fish sampled. *N. unicornis* G36 was used for culturing type Cs and Js, from their respective sections of the gut. *N. tonganus* was used to culture type B cells, the fish were G35 and G77. G78 which did not have type B was used for culturing as well. *A. lineatus* G53 was used for culturing trying to culture type A1 and A2 cells. *A. olivaceous* G49 was used to culture type F, G, and H cells.

Culture medium. Culture media from Table 8 was used, however, only MOPS pH 7.2 and TRIS pH 8.8 buffer were used. Both media types were added to sterile Hungate tubes for 5 mL cultures. They were also added to Hungate tubes containing either 5 mL of solidified 0.5% agar or 0.5% alginate. Type B specific media was also generated and used (Table 9). This was used to inoculate *N. tonganus* G77 gut contents. H₂ gas was added to the headspace of the *A. lineatus* gut content culture in HEPES buffered media.

Table 9 Type B specific media

Component	Amount to add for 100mL (in grams)	Amount of stock added for 100mL	Stock solutions % (w/v)
NaCl	0.8		
KCl	0.04		
CaCl ₂ 2H ₂ O*	0.02	0.2mL	10
NH ₄ NO ₃ ⁻	0.033		
MgSO ₄ 7H ₂ O*	0.1	1mL	10
KH ₂ PO ₄	0.02		
K ₂ HPO ₄	0.015		
FeSO ₄		100μL	4.00E-02

Vitamin K3		50μL	1mg/mL in 100% ethanol
vitamin mix		1mL	
trace mineral mix		1mL	
cysteine HCl	0.06		
resazurin		100μL	0.1
agar	0.001		
L-serine		500μL	
D-glucose	0.01		
D-xylose	0.01		
D-galactose	0.01		
Buffer		10mL MOPS or TRIS	1M
H₂O		86.05mL	

Bold font indicates those that were not autoclaved but were added after filter sterilization. * indicate samples that were autoclaved separately and then added

Samples were checked periodically after inoculation for growth. SYTO9 staining was conducted to stain DNA to check for viability.

B.9 RESULTS AND CONCLUSIONS

Acanthurus olivaceus culturing attempt

Acanthurus olivaceus gut contents were inoculated in both the MOPS and the TRIS buffered media. MOPS was used as a more stable buffer for neutral pHs and TRIS was used due to it appearing to be the best buffer in the previous Lizard Island trip. This fish contained smaller epulo morphotypes like type F, G, and H, some of which undergo binary fission, as well as type As. The tubes became turbid a few hours after inoculation as there was a ton of growth of other bacteria. At 20 hrs post-inoculation, both tubes were investigated microscopically. In most cases the epulos did not appear to look viable. Large type As were destroyed in both media (Figure 14A, B). The smaller epulos present did not appear to progress much in the media and their daughter cells did not look healthy, rather they appeared wilted (Figure 14C, D). There was one small cell type that looked viable (likely type F cells) still in the media, as they had a rough

exterior and did not appear compromised or ghost-like. (Figure 4E, F). However, these cells were not motile or abundant within the cultures, making it unknown if this morphotype was actually viable. By 48hrs post-inoculation, the culture was completely over-run with other bacteria and no longer investigated.

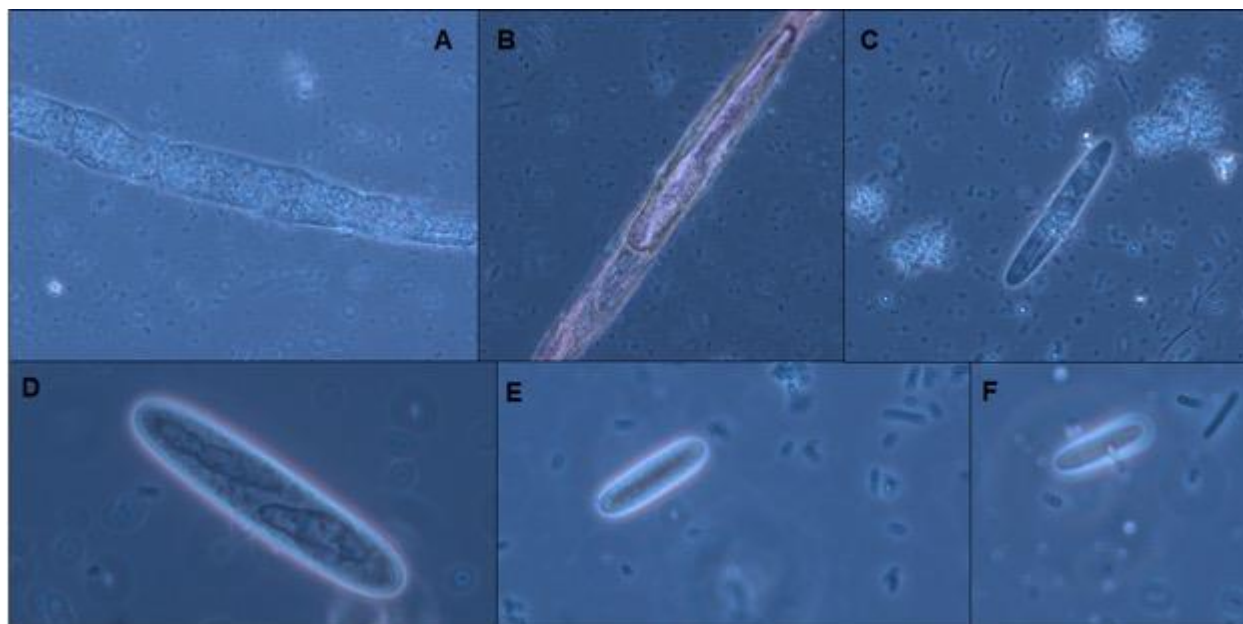


Figure 14 Culture images from *A. olivaceus* gut contents. (A-C) were taken with a 20X objective, while (D-F) were taken with a 40X. (A) and (B) depict type A cells that appear degraded 20hr post-inoculation in the TRIS buffered media. (B) shows a daughter cell within the degraded mother cell, that also appeared degraded. The smaller cells were generally motile, and may have been actively degrading the husks of the type A cells. (C) shows a type H cell in TRIS buffered media, while (D) shows a type H cell in the MOPS buffered media. The cells appeared ghostly and the daughter cells appeared to be irregularly shaped within the mother cell. (E) and (F) depict potentially type F epulos in the MOPS media. Their exteriors appeared rough and almost had a bit of a phase-bright-shine, though they were not endospores. These cells were not motile, but appeared as if they might be viable.

***Acanthurus lineatus* culturing attempt**

Gut contents from *A. lineatus* were used to inoculate MOPS buffered media and H₂ gas was added to the headspace right after inoculation occurred, as type A1s have multiple hydrogenases present in their genomes and may use it for energy conservation. The cells were highly motile in the fish gut contents, but after 4 hours within the media, they were not motile.

They tended to appear phase-bright (Figure 15A-C), while other smaller bacteria were present and motile within the culture. By 12 hrs, most of the large cells appeared degraded and ghost-like (Figure 15D), with a few appearing phase-dark. By 24 hours, most were ghost-like (Figure 15E-F), but there appeared a few phase-dark cells that may have had daughter cell progression stalled at the early stages and may still have been viable in the media (Figure 15G-H). At this stage in time the media was getting overrun by other bacteria, making it unclear if any of these cells progressed further.

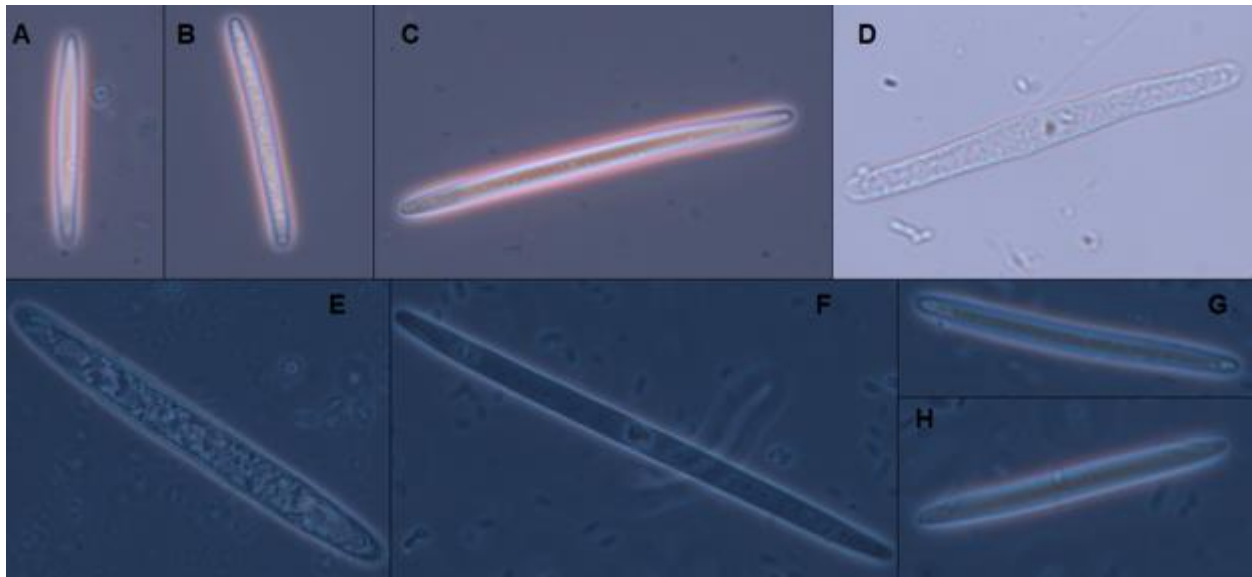


Figure 15 Culturing attempts from *A. lineatus* gut contents. (A-C), (G), and (H) are all from a 20x objective, (D-F) are all from a 40X objective. (A-C) show phase-bright cells at 4 hours post-inoculation. At this point there were not a lot of other bacteria growing in the media, but there were smaller cells present in the culture growing. (D) shows a compromised type A cell, as it no longer had any phase-dark qualities and appeared emptied of its cellular contents. (E) and (F) depict two types of compromised cells present in the media after 24 hours post-inoculation. In (E) the cell appeared textured with intracellular content present, but in disarray. (F) shows an almost completely empty cell with a small disfigured daughter cell at the left pole. (G) and (H) appeared to be more phase-dark. It is unclear if in (G), the cell progressed to this stage, but it is evident at both poles there were 2 daughter cells forming or stalled at this stage of development. What is not completely clear is if the daughter cells appeared more visible because the cell lost its phase-bright nature. (H) daughter cells were visible at the poles, but they were not as phase-bright as in (F). These cells may have been viable, but were not motile and did not appear to be more abundant. Due to the abundance of other things growing in the culture, the tubes were abandoned.

***Naso tonganus* culturing attempts.**

To attempt to culture type B cells, *N. tonganus* gut contents were used to inoculate media. *N. tonganus* G35 contents were used to inoculate TRIS buffered media, and TRIS buffered media also containing agar. Due to their agarolytic genomic potential, we utilized the agar media in hopes pieces of the agar would diffuse into the media, as well as provide type B a surface to attach to and degrade. This was done in the 2-phase culture for 2 reasons, the first being that we could not anaerobically culture onto plates, and the second being that if the agar was degraded by type B, the cells would still be in liquid culture and would not be lost, as what would happen on agar plates. Prior to inoculation, the type B cells were highly motile and contained internal offspring within their cytoplasm. The cells were microscopically checked 22 hours post-inoculation. At this stage, there were no motile cells and most cells did not appear viable in both media tubes. The cells mainly appeared ghost-like, with most of their cell contents not present (Figure 16A, B), but there were a few cells that still had a rough texture and a phase-bright-like coloration (Figure 16C). By 34 hours, there were no viable looking cells, most were glassy and empty of cellular contents (Figure 16E-F). Other cells were growing abundantly in the culture as well. From this it was clear that this media was not successful for their culturing.

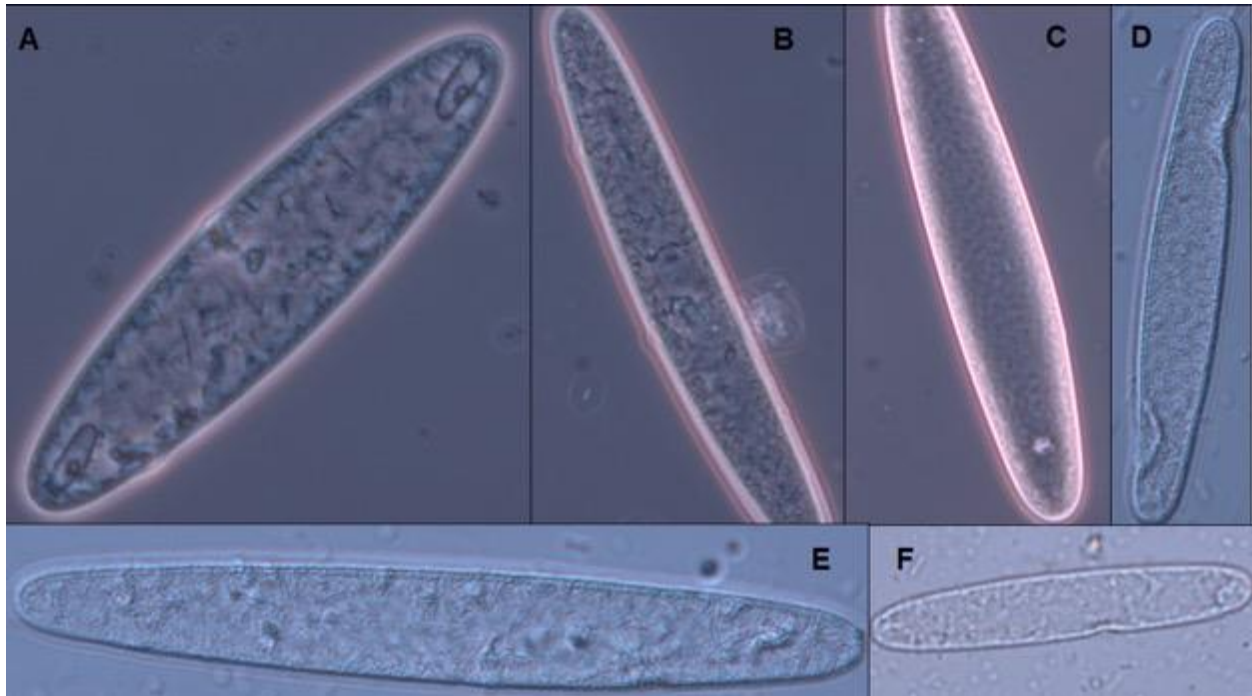


Figure 16 First culturing attempts from *N. tonganus* gut contents. All micrographs were taken with a 20X objective. (A) and (B) are images of compromised cells at 22 hours in the TRIS media without agar. (C) appeared textured and still phase-dark in contrast to (A) and (B), with a daughter cell forming at the bottom pole. This image was taken from the TRIS buffered media as well at 22 hrs. (D) shows a compromised cell with apparent tares in the cellular envelope, at 34 hours in the TRIS buffered medium. (E) and (F) show compromised type B cells in the media with TRIS and agar. There were many small cells around both cells as the media was getting overgrown by other microbiota members.

Due to the lack of promise from the TRIS media, type B tailored media (Table 9) using MOPS as the buffer was inoculated from *N. tonganus* G77. The gut contents of this fish had large type B cells with larger intracellular offspring. At 12 hrs post-inoculation, it appeared as if the daughter cells may have progressed to larger cells, with some appearing viable, but many of the mother cells looked like they were degrading and releasing the daughter cells (Figure 17). *Balantidium jocularum* was present in the culture and was motile. At 16 hours, a few microliters of culture was removed and stained directly with the SYTO9 DNA stain to assess the condition of the DNA and potential viability of the cells. Fluorescence intensity appeared to be highest at the periphery of the mother cells, and the daughter cells typically did not exhibit any high fluorescence patterns, making them overtly discernable from the mother cell, but a few did seem

to be fluorescent and potentially viable (Figure 18). By 24 hours, the cells no longer appeared to be viable and most of the cells present looked deteriorated. *B. jocularum* was still present in the media and was motile. The culture also was quite dense at this time with other bacteria growing. Based on this, it did not appear that the media supported any substantial growth of type B.

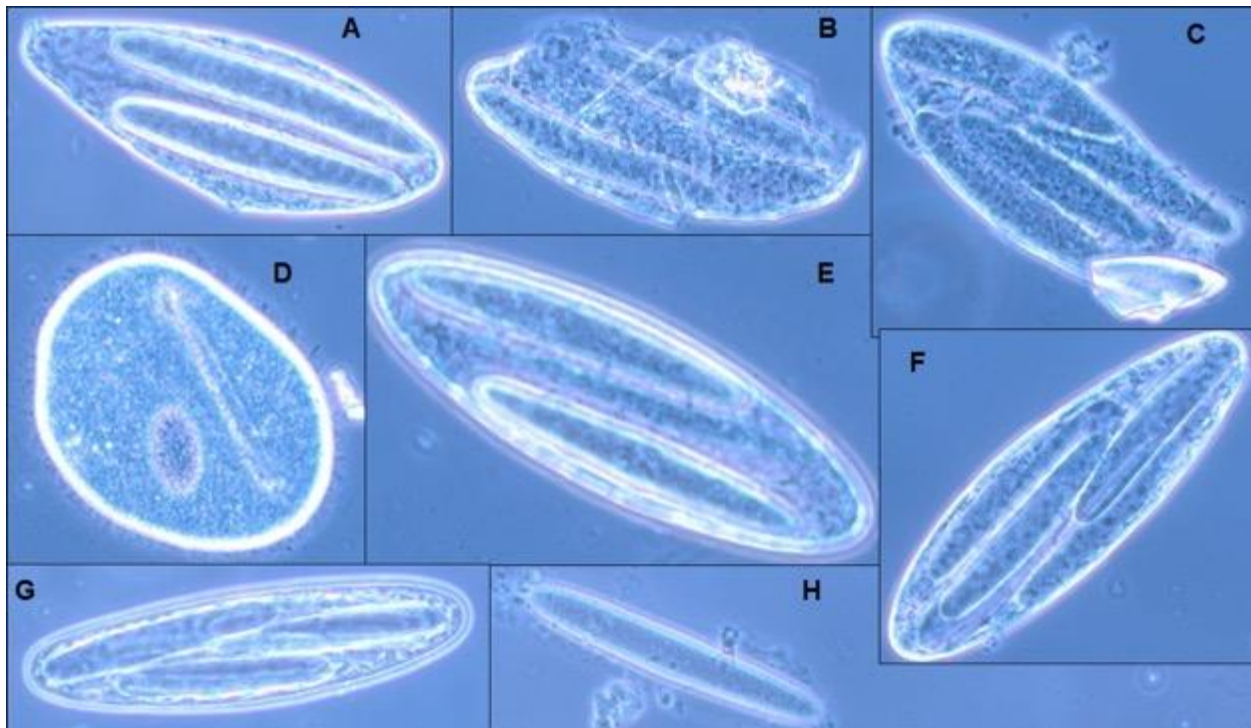


Figure 17 *N. tonganus* G77 type B cells at 12 hrs post-inoculation. All micrographs were taken with a 20X objective. (A-C) depict mother cells with multiple daughter cells present. In these three images, the mother cell does not appear to be in good condition as they were losing shape (A) or appeared to be actively releasing daughter cells (B) and (C). (D) demonstrates a ciliate potentially with a type B cell in its cytoplasm. (E-G) represent mother cells that appear to be in good shape as they maintain their ovoid shape, had some texture to their envelope, and had multiple offspring developing in their cytoplasm. It is unclear if these cells were viable as none were motile, or if the daughter cells had grown larger while in culture. (H) represents a recently released daughter cell, as there appears to still be mother cell debris around it. The cell appears to have texture and not be completely ghost-like or glassy, suggesting it could have been a viable cell released in the media that progressed to a larger cell after release.

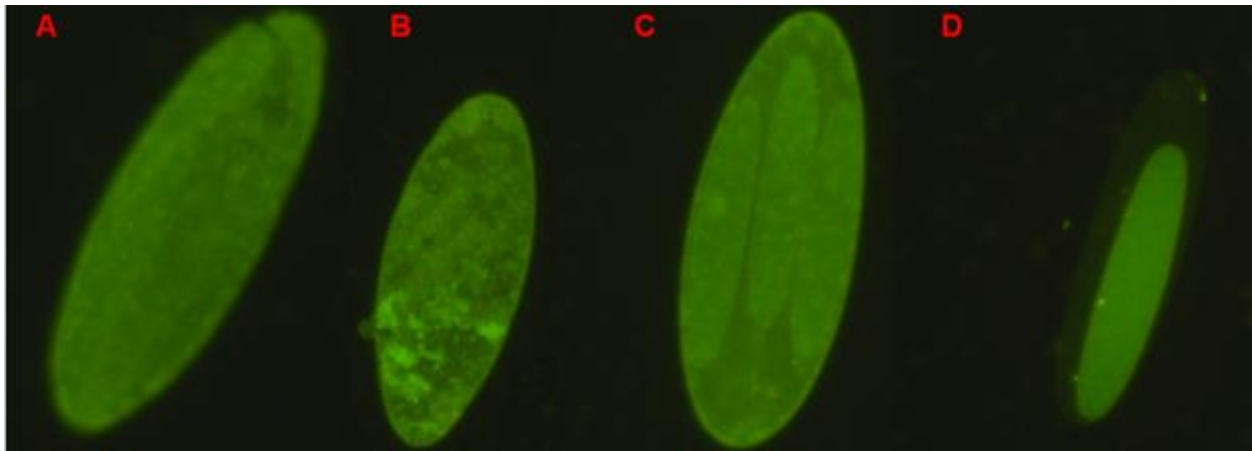


Figure 18 *N. tonganus* G77 type B cells stained with SYTO9 DNA stain at 16 hrs post-inoculation. All images were taken with a 20X objective. In (A) and (B) the brightest fluorescence is visible in the mother cell at the periphery. It is not easy to distinguish the daughter cells based on fluorescence patterns in these two cells. (C) the mother cell was outlined with fluorescence indicating the DNA was located at the periphery. The 3 daughter cells were more distinguishable in this cell, though it was hard to view their staining patterns. (D) exhibits a single daughter cell inside of a degraded mother cell. There did not appear to be any DNA left in the mother cell, but the daughter cell had a uniform staining pattern, making it appear as if it could be viable.

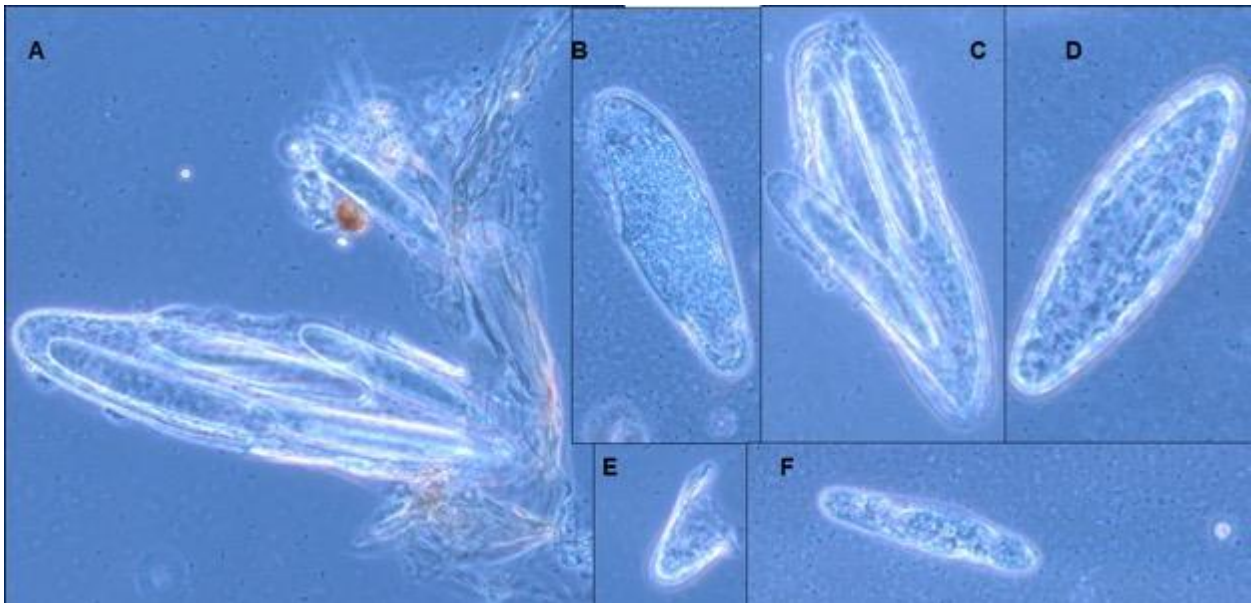


Figure 19 *N. tonganus* G77 gut sample cultures at 24 hr post-inoculation. (A) depicts daughter cells exiting from a mother cell that appeared to be attached to a fragment of some gut content. The daughter cells looked viable in this cell. (B), (D), and (F) all depict degraded type B cells. (B) has a large hole in the envelope either due to release of daughter cells or lysis, while (D) appears to be potentially breaking apart as there appeared to be these bulbous structures at the periphery. (E) shows a piece of a type B cell present in the media, while (F) shows another degraded type B cell, potentially a recently released daughter cell. In (C) the mother cell was degraded but the daughter cells looked okay. The daughter cell emerging appears to be encapsulated potentially within a second mother cell.

N. tonganus G78 was a smaller fish that upon investigation, did not contain any type B cells. Instead it contained smaller epulos, some of them being filamentous, some resembling type C2 cells, and small potential epulos that may undergo binary fission. This was used as inoculum for both the MOPS and the TRIS buffered media, in hopes to get one of these small uncharacterized epulos to grow. The TRIS media appeared to be better for the cells than the MOPS media as more of the MOPS media cells appeared ghost-like in appearance 9-hours after inoculation, while more in the TRIS were larger and appeared textured. Cells in the MOPS media appeared to be phase-dark instead of textured, potentially indicating they they were growing (Figure 20). There was also a longer rod, that resembled spores found in the spore stocks of *N. unicornis* and *N. lituratus*. This was visualized undergoing binary fission within the media (Figure 21). It is unclear if this was an epulo. Cells were then stained with SYTO9 to assess viability. SYTO9 staining revealed that longer cells were uniformly stained (Figure 22). The rougher exterior cells appeared to stain uniformly with the DNA stain, making it possible that some of these cells were still viable within the culture. By 3 days, the culture was completely overran by the other bacteria growing.

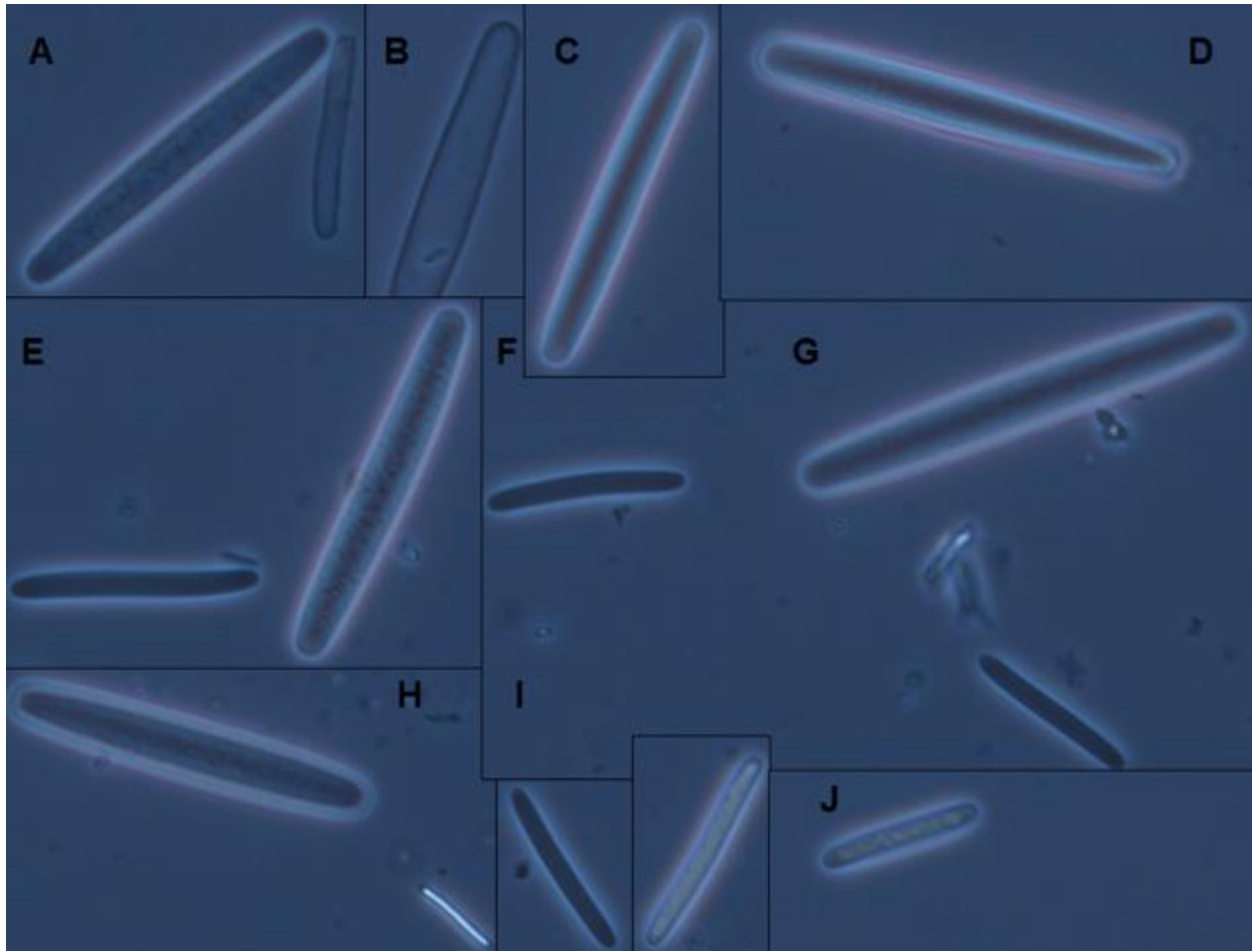


Figure 20 *N. tonganus* G78 cells in MOPS media. All images were taken with a 40X objective. (A) and (B) depict the ghost-like cells that were more prevalent in the MOPS media. The larger cell in (A) appeared to be heading towards the ghostly look. (C-E) and (G-H) show larger type C like cells that appeared textured. These were not as prevalent in the MOPS media in contrast to the TRIS media, but did not appear compromised. The smaller cells in (E-G) and (I) appeared phase-dark, indicating they were possibly viable. They also appear a little darker at both poles, possibly indicative of daughter cell formation within the media. (H) also depicts a small phase-bright endospore that looked similar to the viable cells found undergoing binary fission in the TRIS buffer. (I) and (J) contain smaller epulos that seem to be the same morphotype as the phase-dark cells, however, these are textured and have a phase-bright nature. The cell in (I) may have been forming daughter cells at the poles.

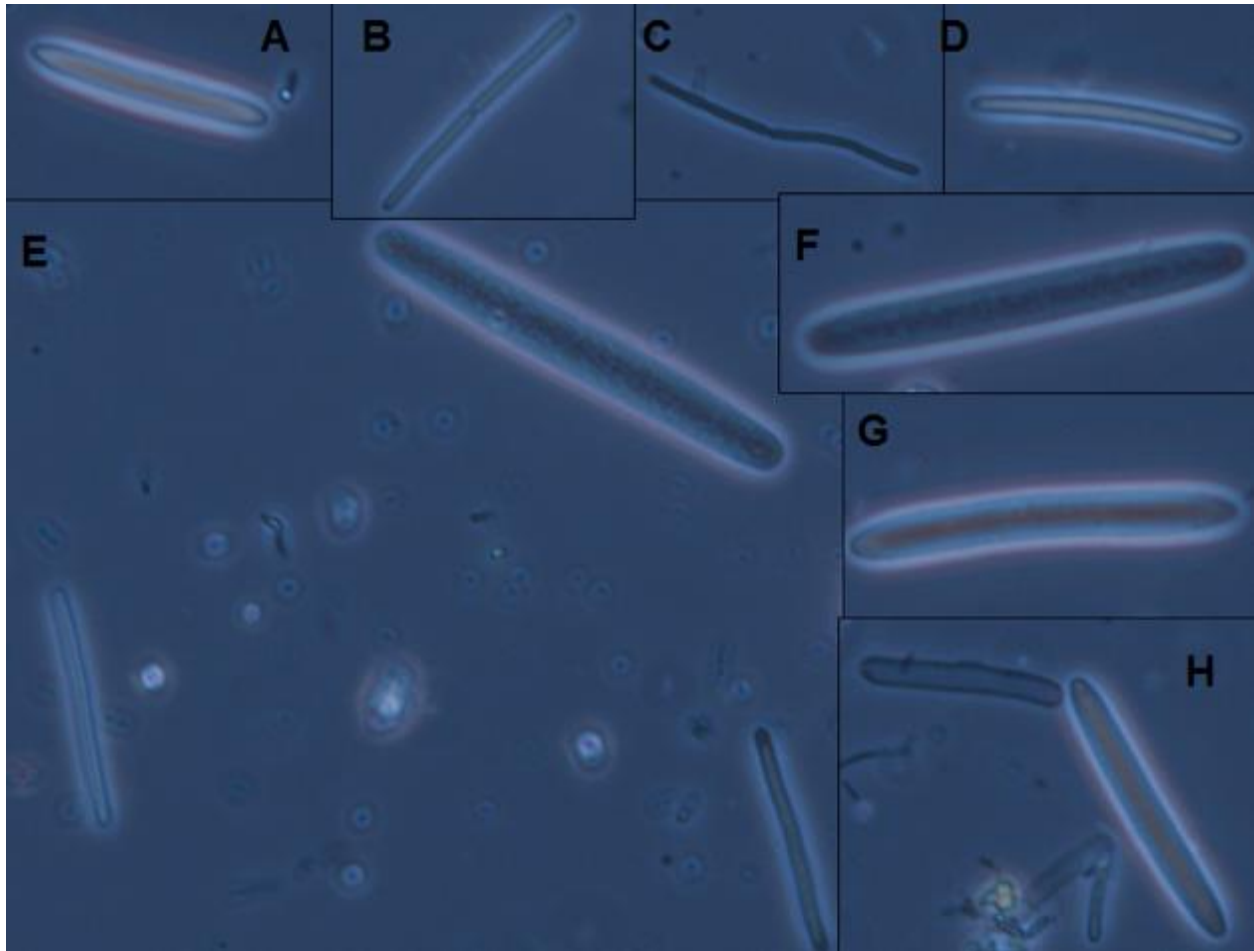


Figure 21 *N. tonganus* G78 cells in TRIS media. All images were taken with a 40X objective. (A) depicts a smaller epulo with a rough texture that had a phase-bright quality and rough texture. (B) and (C) might be an epulos, as they resembled vegetative forms of single spores found in previous *N. unicornis* and *N. lituratus* spore stocks, but this was not clear. Both of them were shown undergoing binary fission. (D) shows a thin, needle-like epulo with a phase-bright tint to it. (E) depicts 2 more of these needle-like epulos, the one to the bottom right appeared to be becoming phase-bright as there was a phase-bright outline appearing inside the mother cell. Also in (E) was a larger cell with a rough texture, resembling type C cells. This cell type is further depicted in (F) as a similar sized and textured cell is present. Both do not appear to be compromised, and could have been viable. (H) resembled the larger cells in (E) and (F) but has more of a shine and rigidity to it. (I) depicts a ghost-like epulo cell next to another of these rigid, almost phase-bright cells.

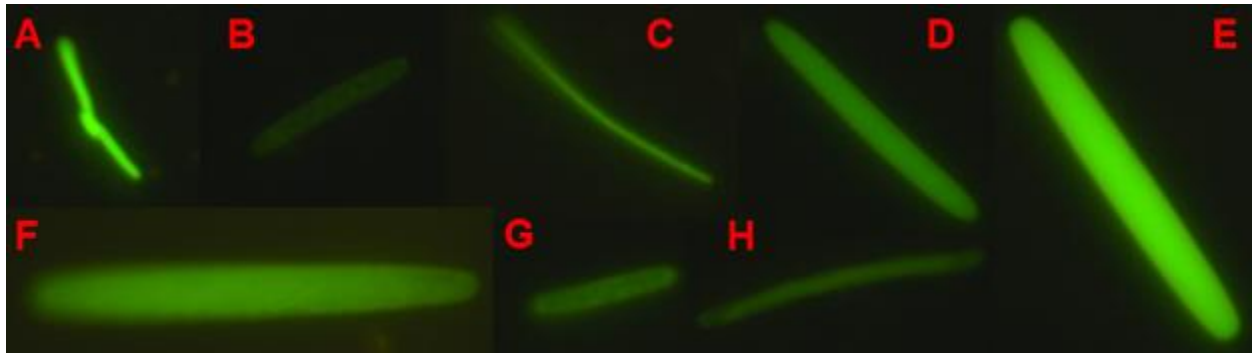


Figure 22 *N. tonganus* G78 cells in MOPS media at 10hrs stained with SYTO9. All images were taken with a 40X objective objective. (A) depicts one of the potential epulos undergoing binary fission, with uniform staining throughout the cells. (B) shows a smaller epulo that was phase-dark, with DNA appearing to be coalesced at the poles. (C) depicts a long potential epulo with uniform DNA staining throughout. (D-F) depict the larger epulos found in the gut. They had uniform staining throughout. The cells had a rough appearance so it was not clear if the uniform staining was because of a quality of the cell's surfaces or if it was bright staining throughout the cell because of DNA. (G) shows patchy patterns of DNA staining throughout the periphery of the cell. (H) did not have particularly bright staining, so it was unclear if the cells are viable or not.

***Naso unicornis* culturing**

N. unicornis culturing was done with insight based on the previous culturing attempts at Lizard Island. Culturing was attempted from both the type C and J rich sections of *N. unicornis* guts. These were used to inoculate both TRIS and MOPS buffered media as well as TRIS buffered media with alginate. In all the cultures, the cells originally appeared viable, though no motility was observed in any of the cultures. Binary fission was witnessed for multiple morphotypes in the media. The TRIS buffered media inoculated from the type C heavy region appeared to allow for progression of the cells through life stages, as it appeared that smaller type C cells underwent a cycle of sporulation (Figures 22-25). The MOPS media inoculated from the type C section appears to have also shown progression of type C cells including some of the larger cells. Both these cultures were eventually overrun by other bacteria. The TRIS media with alginate appeared to potentially be maintaining type J cells, and appeared to only contain epulos and epulo-like cells. Binary fission was still viewed within the culture 6 days after inoculation, suggesting that the type J cells were being maintained in the culture. At certain points, it

appeared as if certain types of J-like cells were increasing in numbers, though there was no quantitative data to support this. Interestingly, these J cells that appeared to be maintained were inoculated from the type C heavy section. Unfortunately, frozen stocks of these cultures were destroyed in transit back from Lizard Island. Quantitative data was not acquired, but this culture had promise and may have had cells living at a slower, non-motile rate.

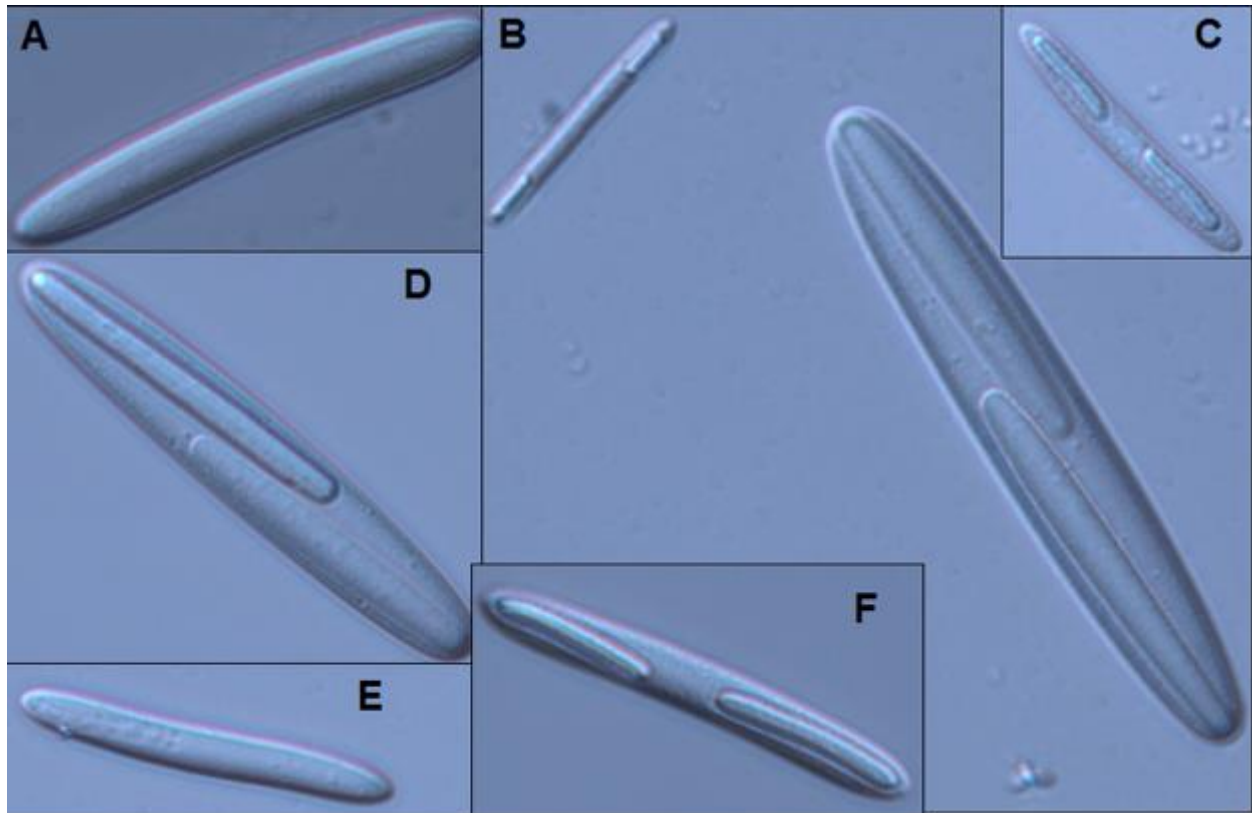


Figure 23 TRIS buffered media inoculated from the type C region of *N. unicornis* G36 at 6hrs. All images were taken with a 40X objective lens. The figures show the diversity of epulos found in the media. (A) Shows a healthy looking type C with offspring developing at the poles. (B) depicts a small type C with spores at both poles, as well as a large type C that had large spores present, this same cell type was represented in (D). (C) shows another smaller epulo that had spores present, this is also demonstrated in (F). (E) shows an epulo with a rougher exterior and some texturing at the left end pole, where a daughter cell appeared to be developing.

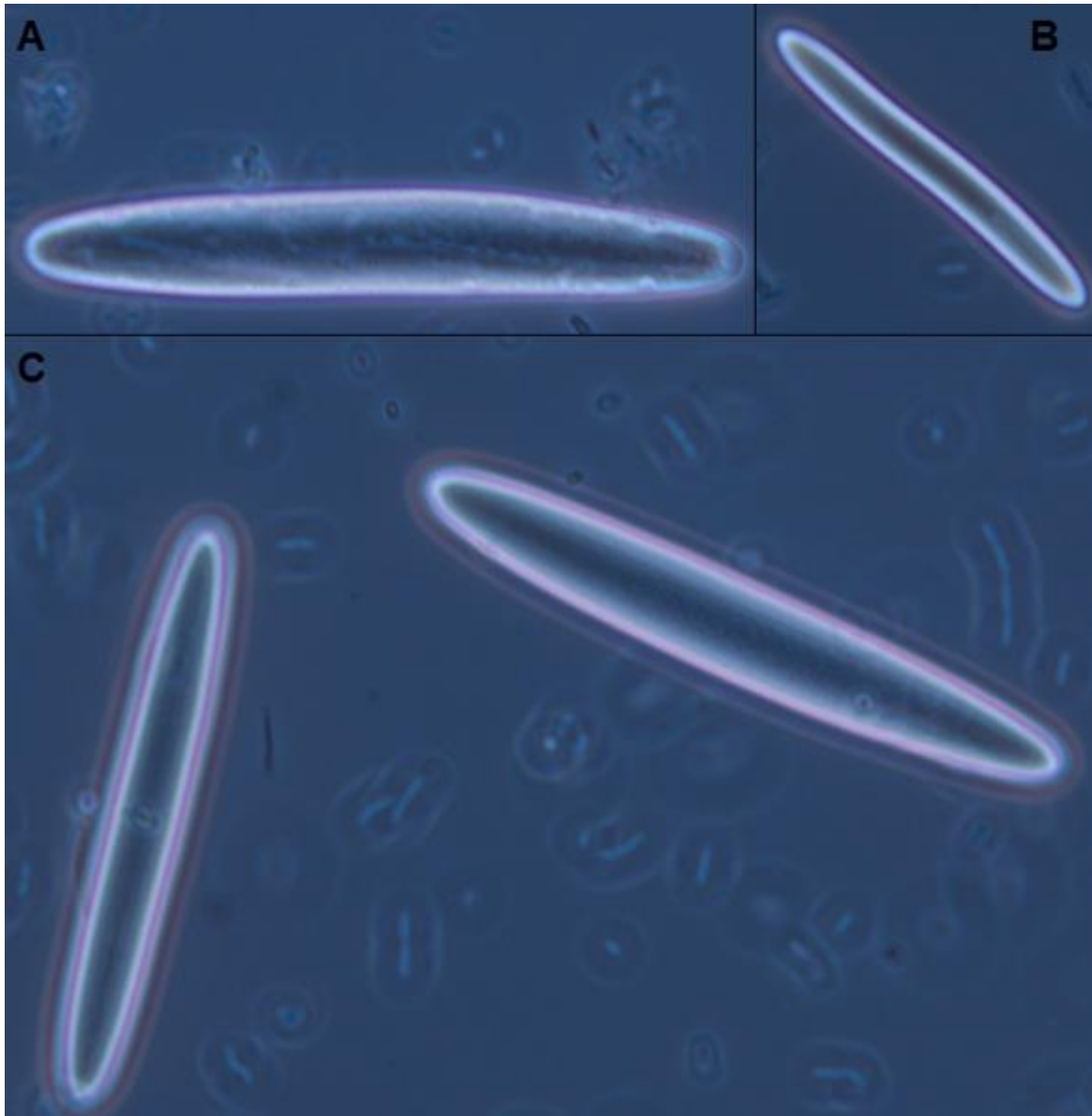


Figure 24 depicts the progression of the TRIS buffered media inoculated with type C's from G36. All images were taken with a 40X objective lens. (A) and (B) show type C cells at 20 hrs post-inoculation, and (C) shows them at 30 hrs post-inoculation. The cells in (B) and (C) appeared to be in good shape and did not appear to have very large daughter cells present. The cells in (C) appeared as if they were recently released from mother cells and may have been viable.

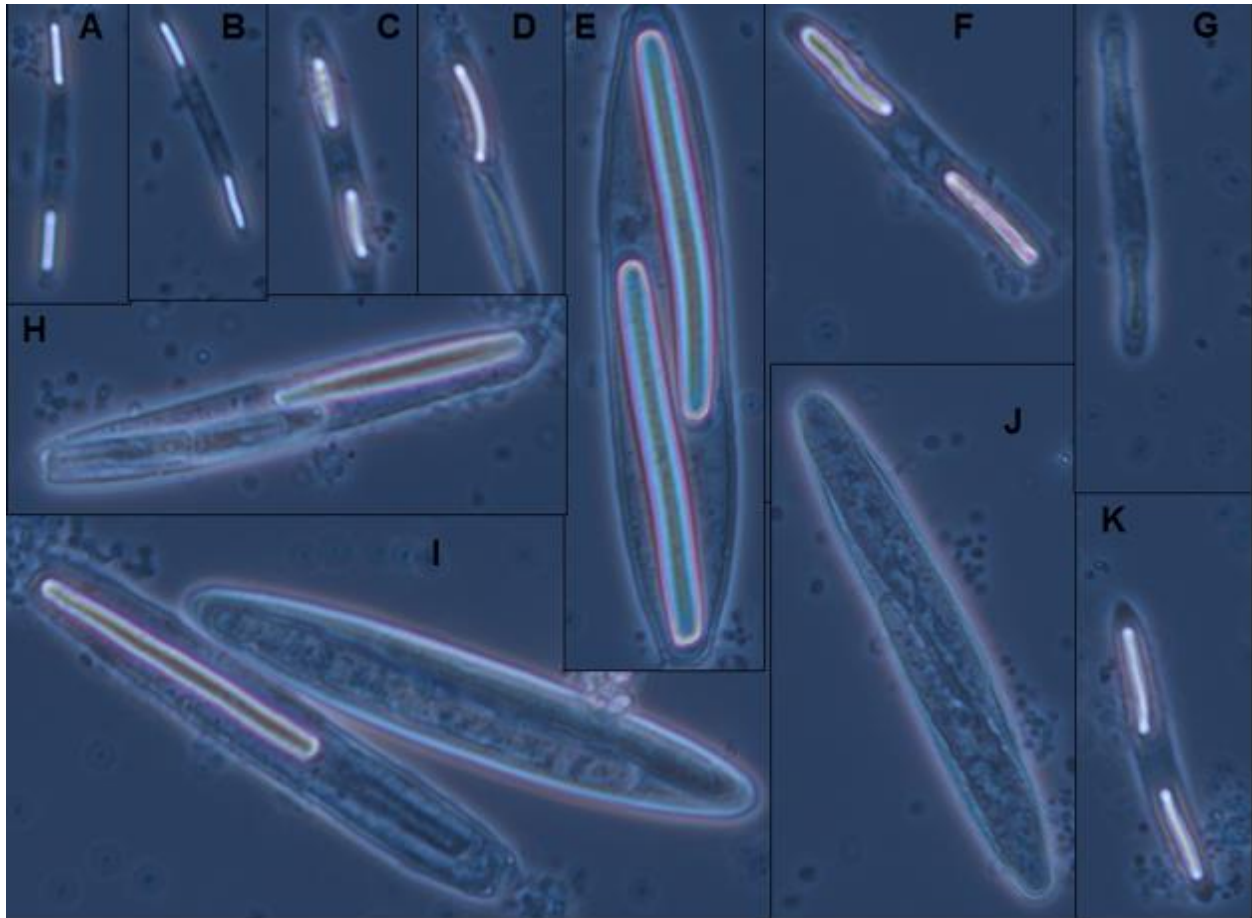


Figure 25 The progression of the TRIS buffered media 120 hrs post-inoculation from the type C area of G36. All images were taken with a 40X objective lens. The figure shows certain cells appeared to progress into sporulation in the media while others appeared to die. The smaller type C's represented by (A-D), (F), (G), and (K), depict the different outcomes. (A-C, K) all appear to have progressed to sporulation in the media, with properly shaped phase-bright endospores present in a degraded mother cells. (D) depicts a cell with one phase-bright endospore and another endospore that either did not form properly or had yet to fully mature. (E) shows a large type C with forespores not yet fully phase-bright, suggesting they were still developing in the mother cell. (F) shows phase-bright spores that appeared to be misshapen, thus they may have been developmentally defective. (G) shows another small type C that died and degraded before it could form mature endospores. (H-J) show large type C's with daughter cells of different outcomes. It appears that the cell in (H) and the bottom cell in (I) both have one daughter cell that was developing into a phase-bright endospore, while the other daughter cell was degrading. In both of these cells the mother cell was degraded. The top cell in (I) appeared to be in better shape as it still maintained its shape, however its daughter cells looked degraded. (J) shows a mother cell that was in poor shape with degraded internal offspring.

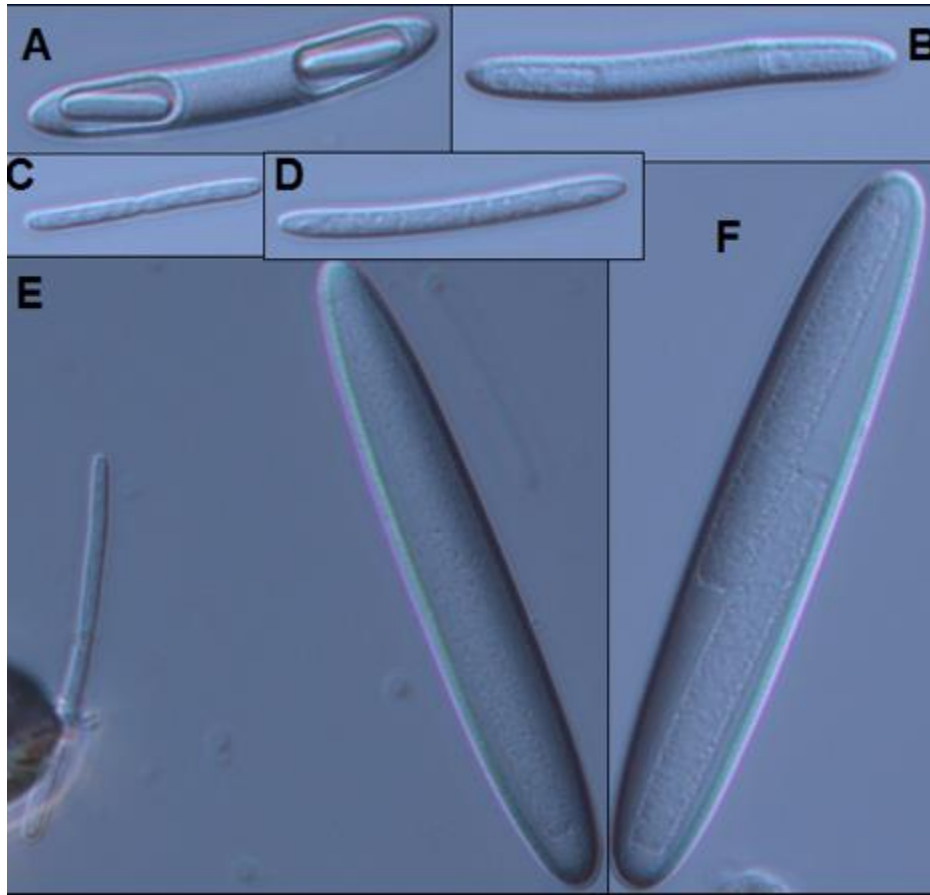


Figure 26 shows the MOPS buffered media inoculated with cells from the type C section of the gut of *N. unicornis* G36 at 6 hrs post-inoculation. All images were taken with a 40X objective lens. (A) and (B) show smaller type C cells with different looking endospores present at the poles. (A) appears to have these deep pockets containing the endospores. (C) shows a small epulo-like cell undergoing binary fission in the media. It appears as if there may have been endospores present in each cell's poles. A similar cell type was shown undergoing binary fission in (E) as well. (D) shows a C2 cell with a rough texture and endospores developing at the poles. (E) and (F) show large type C cells with distinct daughter cells present.

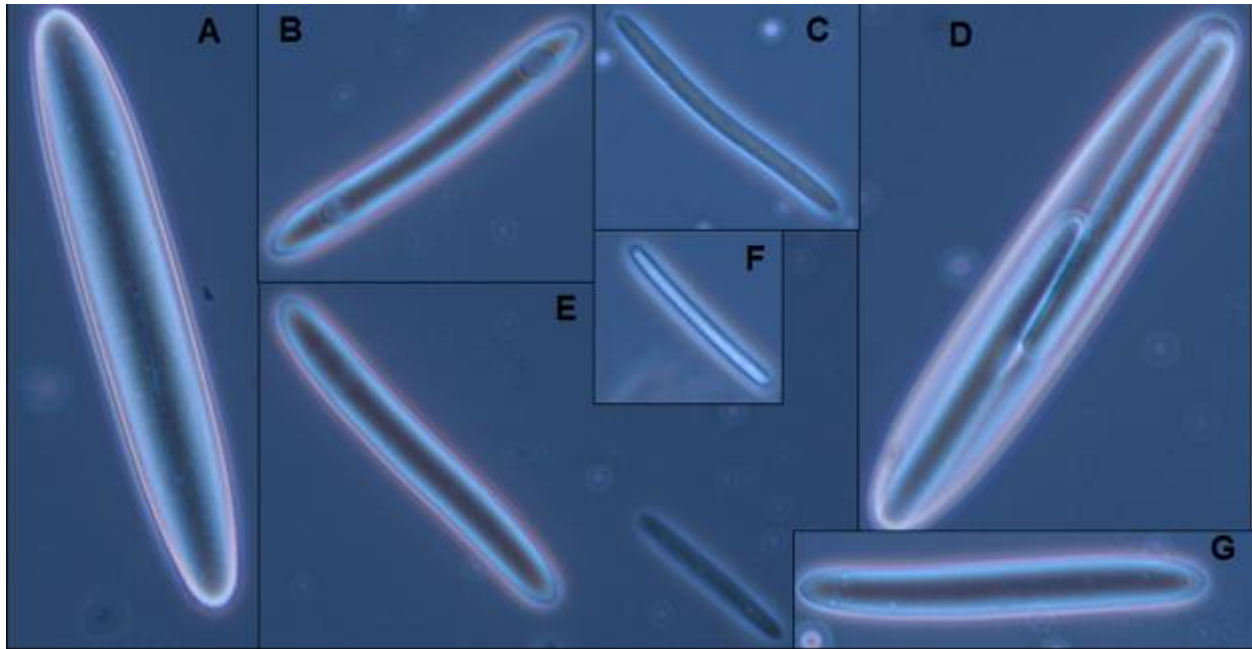


Figure 27 depicts the progression of the MOPS buffered culture from the type C section of *N. unicornis* G36 at 20 hrs post-inoculation. All images were taken with a 40X objective lens. (A) a large type C cell that appeared textured with endospores developing within. The cell appeared to be in good shape and looked like it may have developed from an emerged endospore within the media. (B) shows a smaller C2 with endospores developing at the poles. (E) and (G) show similar cell types without discernable daughter cells developing. (C) shows an epulo with small daughter cells that appeared at the poles. This was the same morphotype that underwent binary fission at 6 hrs. (D) shows a large type C that seemed to be progressing further than the cell in (A). The endospores were more distinct looking in the mother cell cytoplasm and were beginning to take on a phase-bright shine at their edges, suggesting that this cell was possibly progressing within the media. (F) depicts a phase-bright endospore that likely developed in the media.

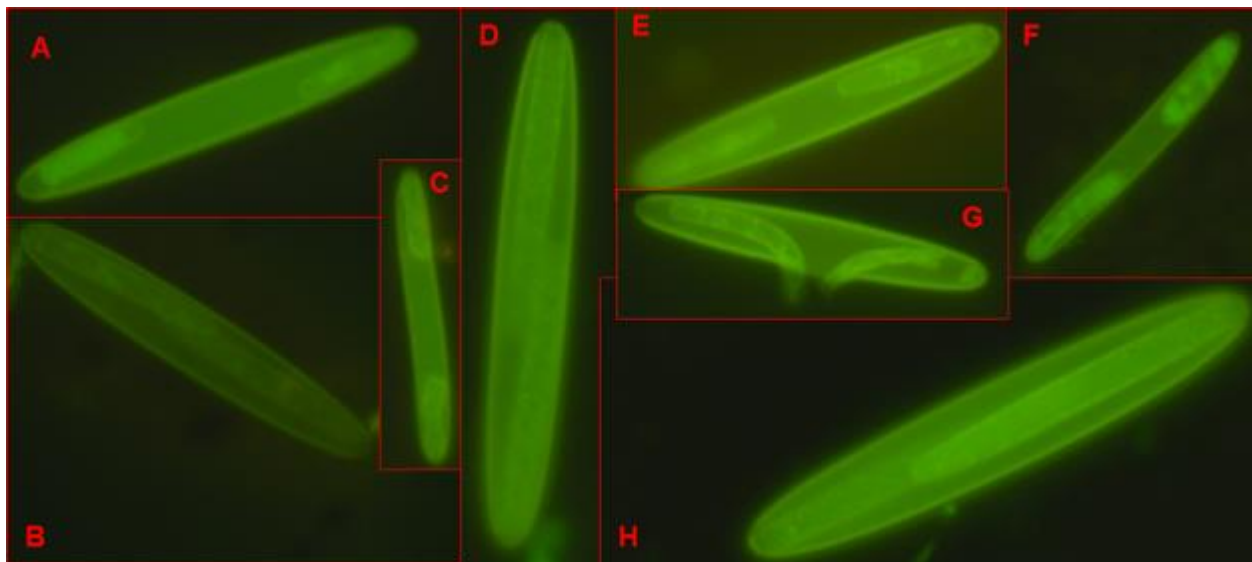


Figure 28 The progression of the MOPS culture inoculated with G36 *N. unicornis*, stained with SYTO9. All images were taken with a 40X objective lens. (A-C) show cells stained at 24 hrs

post-inoculation, while (D-H) show cells stained at 30 hrs post-inoculation. This figure shows good staining of the mother cell peripheries suggesting that they were still viable and not degraded when the cells were imaged. Strong staining of the mother cell DNA can be seen in all figures. The endospores showed different degrees of staining. (A) the spore on the left appeared to be almost uniformly stained with SYTO9 except for at the right pole, while the right pole spore only had a spot of staining. The spores in (B) exhibited a little bit of staining at their peripheries, as with those in (C). Stronger daughter cell staining was viewed in (D) and (H) as the spore peripheries were stained, and there appeared to be blotches of staining throughout these daughter cells. This suggests that the cells in both panels were viable, in good condition, and the forespores were developing properly. (E-G) show splotchy staining patterns of the spores. The daughter cells in (F) appeared misshapen and as if they were releasing from the mother cell.

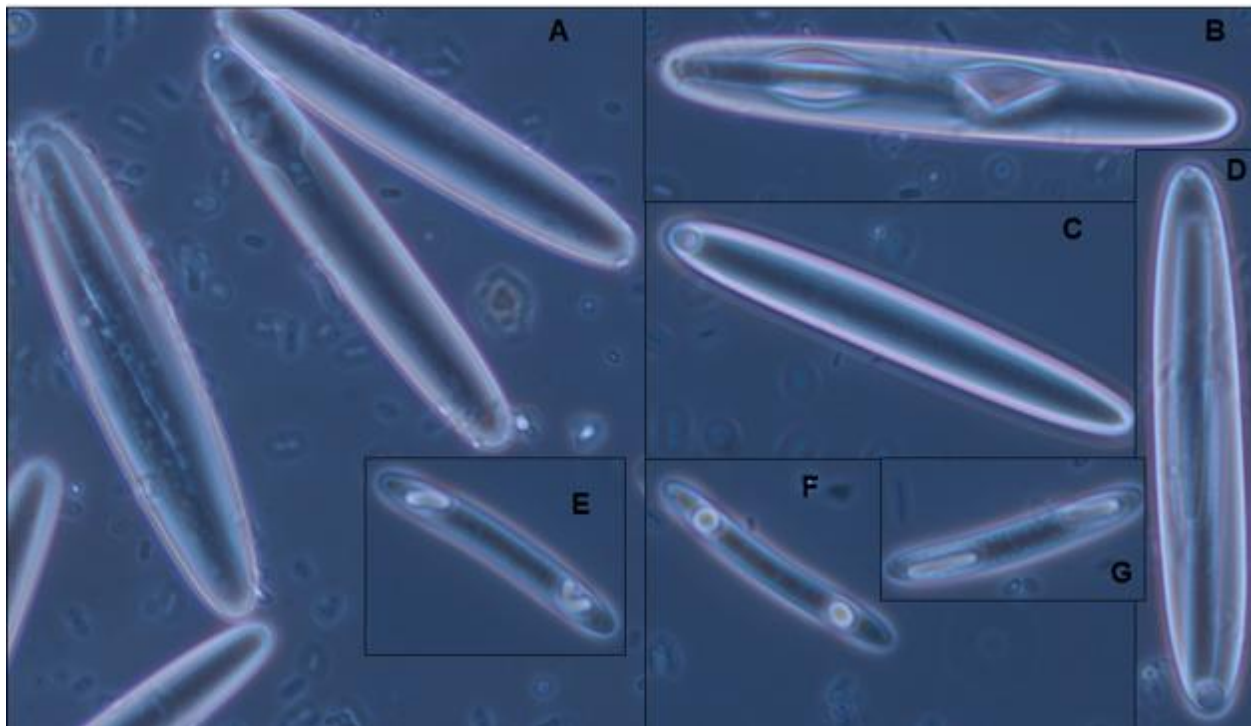


Figure 29 The progression of the MOPS culture inoculated with G36 *N. unicornis*, at 30 hrs post-inoculation. All cells imaged with a 40X objective lens. The culture at this stage was being overrun by other bacteria, as exhibited by all the small cells present in the images. (A) shows multiple large type C cells, where in each case the mother cell appeared to be intact, with daughter cells developing at different stages. The cell to left appeared to have intact daughter cells developing, the central cell appeared as if the daughter cells were lysed within the mother cell, and in the top mother cell, it was hard to discern the status of the daughter cells. (B) shows daughter cells being released from the mother cell. These were not phase-bright, so it could have been an abortive process or that they germinated in the mother cell. (C) and (D) show large type C cells. Both cells show a large circular structure at the pole of the cell, possibly a new daughter cell. (E-G) show smaller type C cells with different degrees of endospores forming. In (E) the forespores appeared to be either misshapen or semi-formed, as one was curled. In (F) the forespores were still circular and not yet rod shaped as typically observed in these cells. (G) shows forespores that had the more typical spore shape.



Figure 30 TRIS buffered media with alginate inoculated from the type C section of G36 *N. unicornis* guts at 20 hrs post-inoculation. All images were taken with a 40X objective lens. (A) and (B) show long type J cells that were undergoing binary fission within the media. This suggests that the cells were able to survive in the media and divide within 20 hours post-inoculation. (C) shows a smaller epulo with daughter cells developing at the poles. (D) and (E) show degraded larger type C cells. Most of the large type C cells were degraded at this time in the culture. (F) shows a healthier larger type C cell with forespores developing at the poles. (G) shows a phase-dark type J cell. (H) shows a similar morphotype to what was shown in (C), though the cell appeared a bit more phase-bright. (I) shows a potential type J cell.

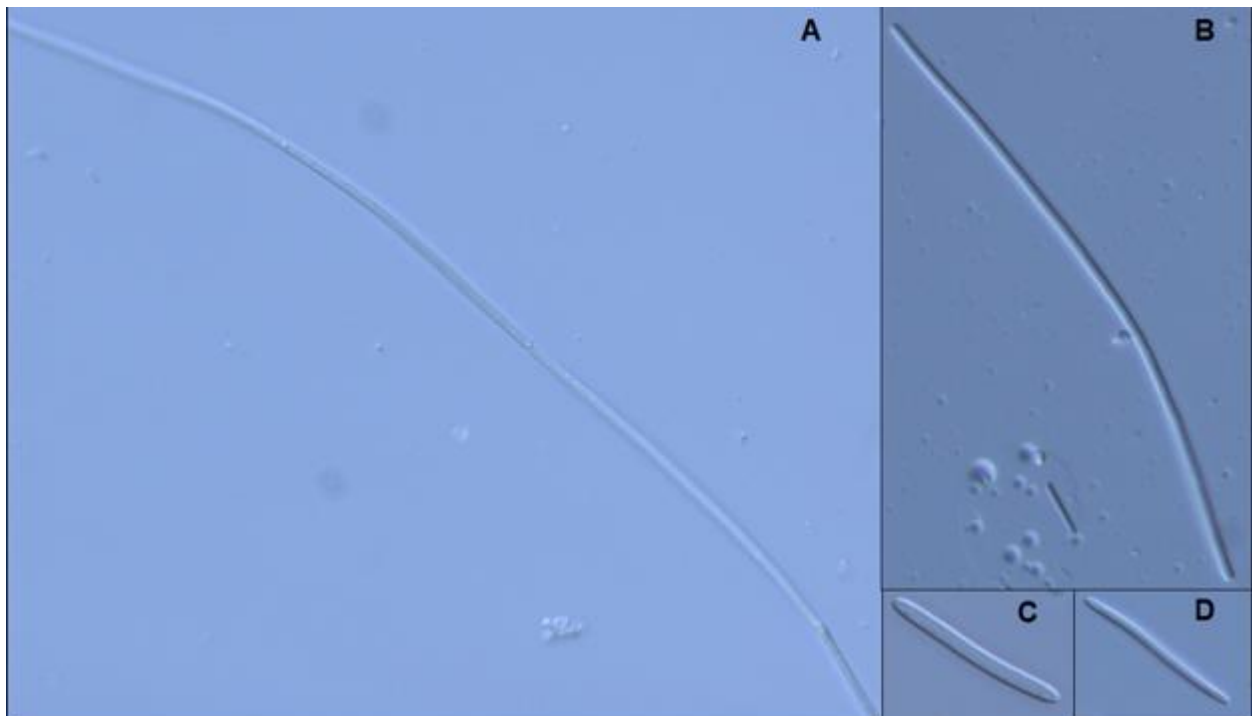


Figure 31 TRIS buffered media with alginate inoculated from the type C section of G36 *N. unicornis* guts at 38 hrs post-inoculation. All images were taken with a 40X objective lens. (A)

and (B) show larger type J cells undergoing binary fission, suggesting they are still metabolically active in the culture. (C) and (D) show smaller epulos with daughter cells developing at their poles suggesting they were still active in the culture.

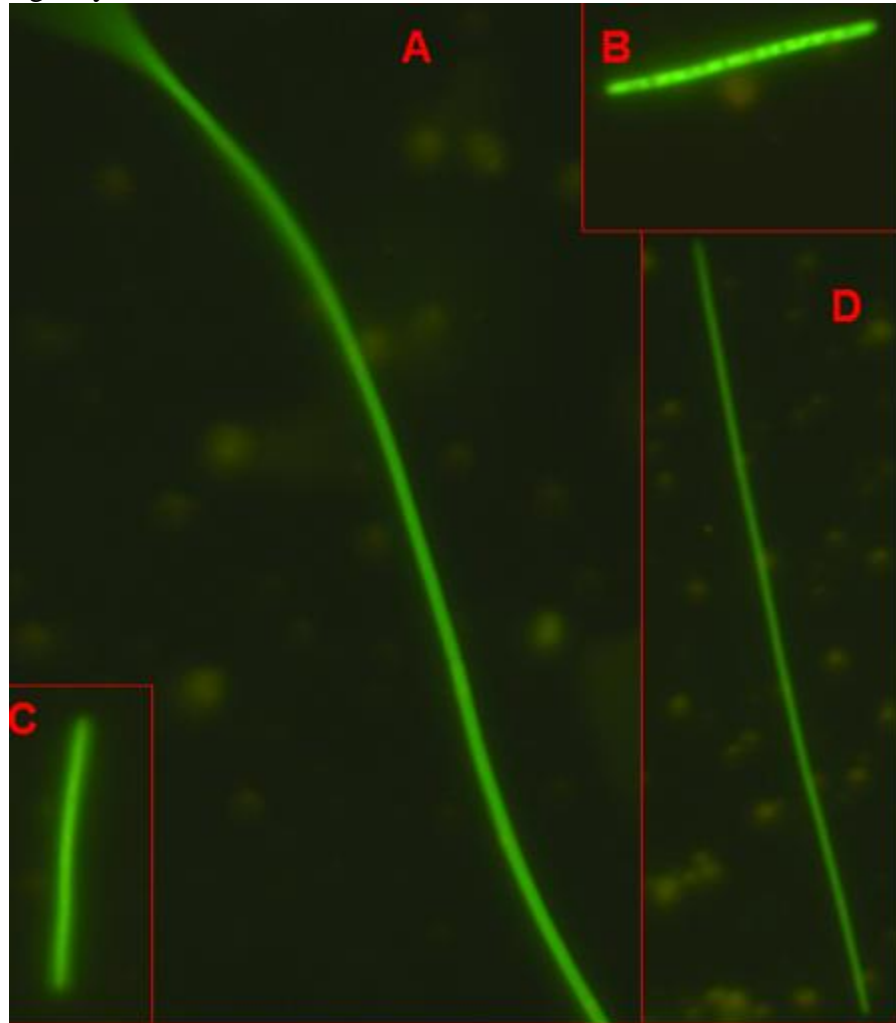


Figure 32 TRIS buffered media with alginate inoculated from the type C section of G36 *N. unicornis* guts at 52 hrs post-inoculation, stained with SYTO9. All images were taken with a 40X objective lens. The type J cells showed strong homogenous staining throughout their cells and showed some cells dividing. (A) shows a dividing large type J cell. (B) shows more of a splotchy pattern of DNA staining. (C) shows a similar looking cell that had uniform staining throughout. (D) shows a thinner type J cell with DNA staining throughout. The strong DNA staining in these cells and the fact that some are dividing suggests that the cells were viable and metabolically active within the culture.

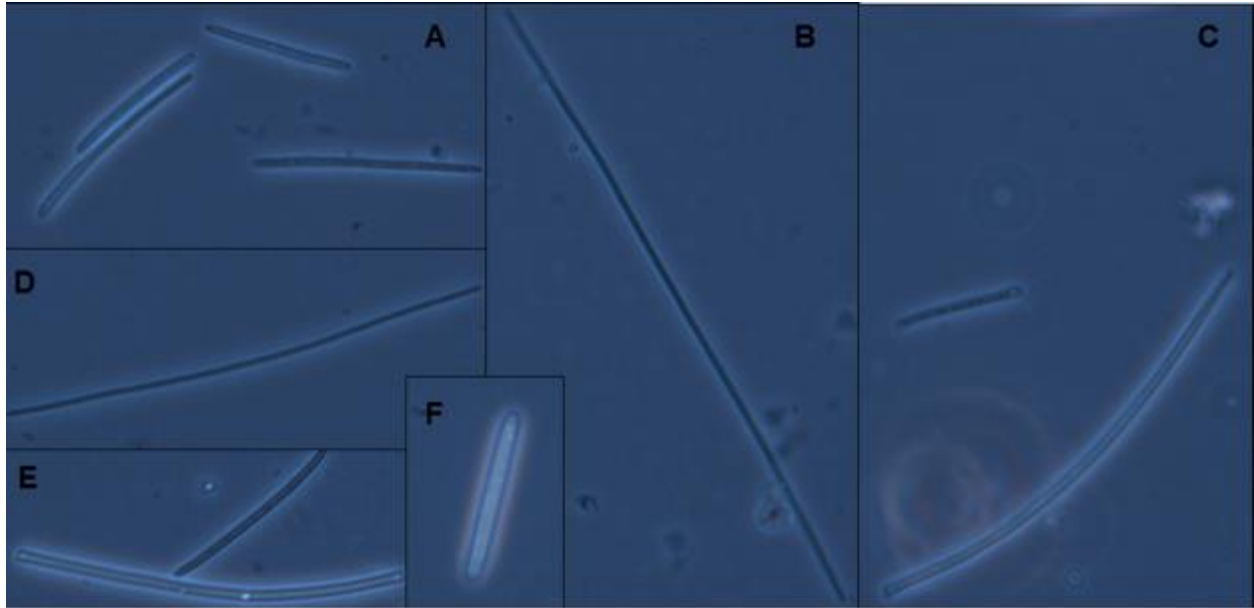


Figure 33 TRIS buffered media with alginate inoculated from the type C section of G36 *N. unicornis* guts at 52 hrs post-inoculation. All images were taken with a 40X objective lens. At this stage of the culture, the cultures appeared to mainly only contain epulos with very little growth of any other bacteria. The cells in (A) and (E) were indicative of what was common in the culture. (B) and (D) show phase-dark long type J cells. (C) shows a type J cell with a pointed end. The cell in (F) was not very common in the media, but had a phase-bright tint to it. These cells looked good, suggesting that they were still viable in the culture, and were possibly growing, though no motility was observed.

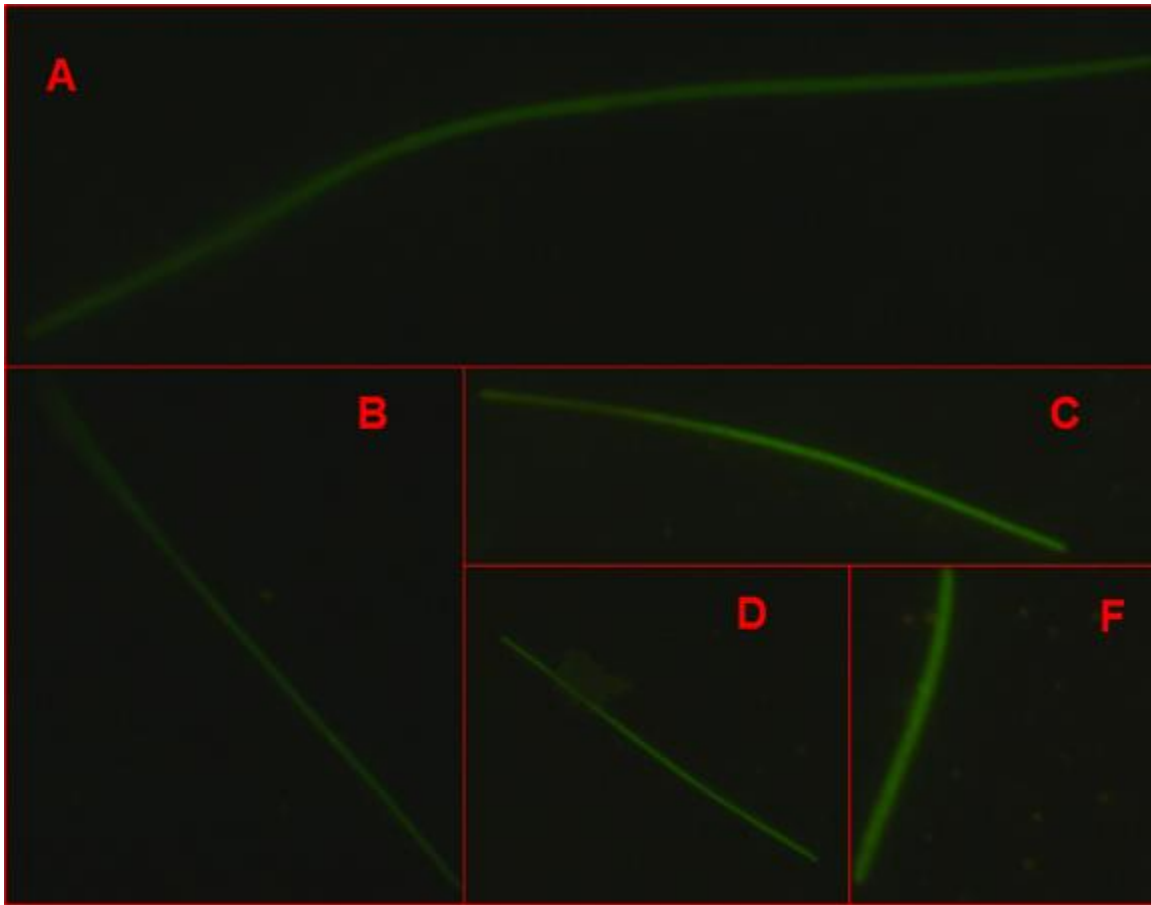


Figure 34 TRIS buffered media with alginate inoculated from the type C section of G36 *N. unicornis* guts at 110 hrs post-inoculation, stained with SYTO9. All images were taken with a 40X objective. SYTO9 staining revealed homogenous staining in the type J cells, suggesting their DNA was still intact and the cells were still potentially viable.

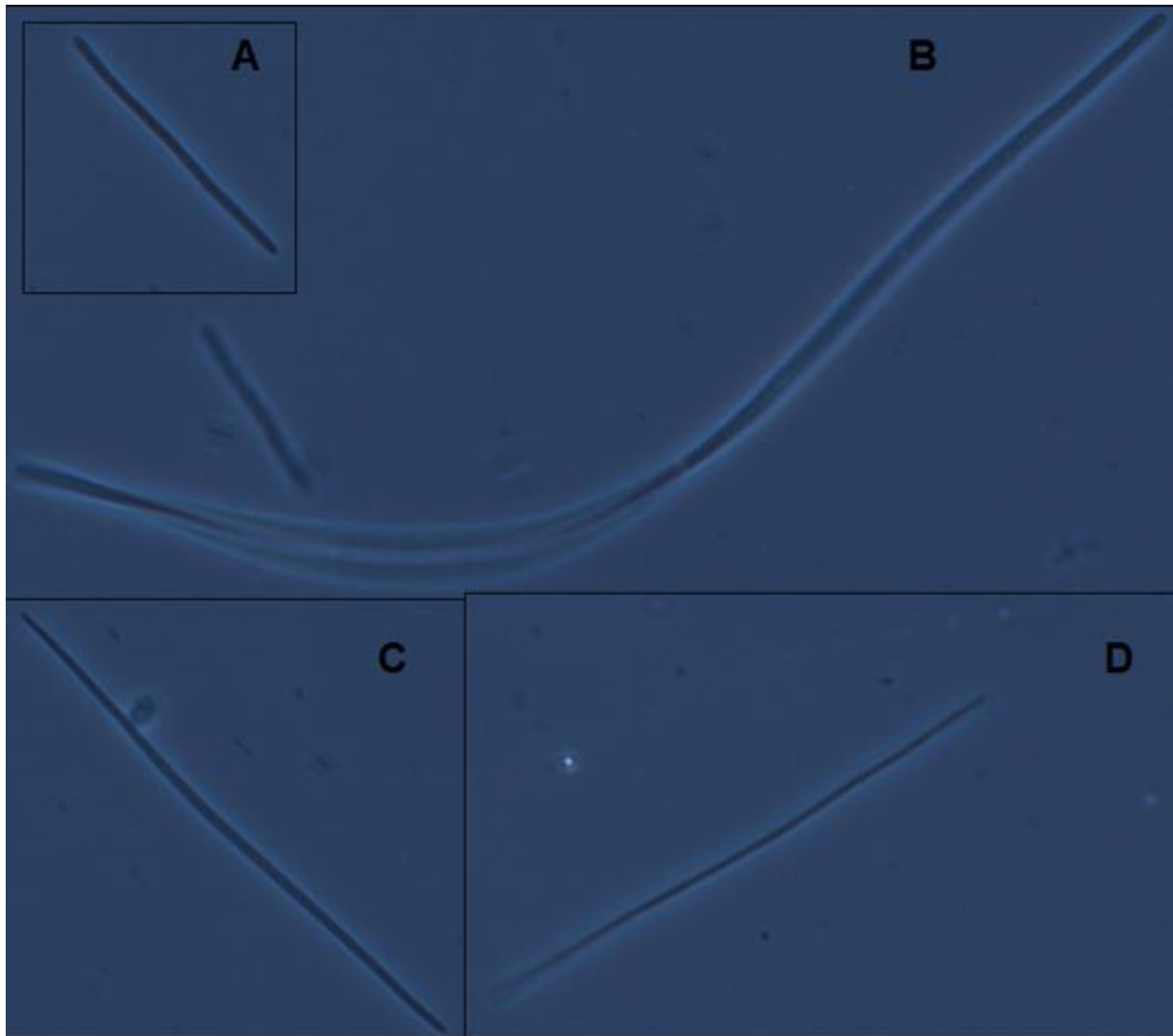


Figure 35 TRIS buffered media with alginate inoculated from the type C section of G36 *N. unicornis* guts at 144 hrs post-inoculation. All images were taken with a 40X objective lens. The figure depicts different type J cells that are still present in the media. All of which appeared phased dark. (B) depicts a larger type J undergoing binary fission. The phase-dark nature and presence of binary fission suggests that the cells were still metabolically active or viable in the culture.

Based on the live culturing directly from gut contents, there were some common themes found. The really large morphotypes did not survive very well in their respective culture media as the type A's did not survive for long. The type B cells fared a little better as some may have been in better shape prior to lysing, but ultimately did not survive. The larger type C cells appeared to fare better in all the medias they were tested in. TRIS buffered media at pH 8.8

appeared to allow for more endosporulation of smaller morphotypes, while the MOPS media appeared to allow for potentially more progression in the larger type C cells. Alginate appeared to spark the survival of type J and J-like epulos as they were the main types of bacteria found in the TRIS buffered media containing alginate. They were undergoing binary fission 6 days post-inoculation and had uniform DNA staining indicative of viable cells, suggesting that some background growth may have been occurring. This proposes that the cells were able to persist, but the media is not ideal for growth. The media recipe could serve as a base for maintaining cells, but modifications need to be made to enhance potential culturing in the future.

B.10 DISCUSSION

None of the culturing attempts were clearly successful at allowing for rapid growth of epulos. Certain conclusions can be drawn from these culturing attempts. Using spore stocks, although it allowed for culturing attempts at Cornell were not ideal. This is because there are multiple other spores present in the stocks, and these all got mixed together based on the spore collection procedure from gut contents. This complicates the studies as a lot of other bacteria are able to germinate and grow in the media. For the future, it would likely make sense to not spin down and concentrate the spore stocks from the gut, rather try to collect directly from the most abundant type C and type J sections of the gut. This could potentially reduce some of the other bacteria from growing and over-growing the cultures. It is unclear if their presence has a negative effect on the growth of epulos, but controlling for these factors could improve results. The other large problem with using endospores is that spores have to germinate and then grow. Although a general baseline for the endosporulation status of the stocks can be assessed, it is difficult to truly know the status of all the spores. It is likely that some of the germination did

actually occur in the media, however, it cannot be excluded that some of the cells that appeared to germinate were already germinated in the stocks. This could conflate the numbers of germinated spores. Regardless, there did appear to be an increase in the number of phase-dark cells in the cultures. Motility is a good measure for metabolic activity, and unfortunately, no motility was observed in any culturing attempts. Using spores makes this even more difficult, because there are no motile cells present at the beginning of culturing from a spore stock, so it could not be observed if the media stopped allowing for metabolic activity.

Using live cells appeared to produce the best results. This was particularly the case with type J cells as they appeared to be undergoing binary fission within the TRIS buffered media containing alginate for almost a week post-inoculation. Although we did not observe any motile epulos in our media, going from live motile cells, atleast provides a proxy for their metabolic status. The fact that the media was unable to support motility demonstrated that specific conditions of the media were not ideal. It could be that not enough, or not the right substrates for energy conservation were provided for motility of the cells, or there were other conditions, possibly pH or Na^+ concentration that did not allow for motility. Having the motile inoculum provided this insight, something that using spore stocks lacks. Due to the general synchronicity of the populations of cells in the fish gut, it also allowed for us to determine if life cycles appeared to progress within the media. For trying to culture type C and J cells, it appeared as if using the higher pH media worked better than more neutral pH media, but due to the lack of definitive growth this remains unclear.

Now with a better understanding of the metabolic capabilities of certain epulos it might be possible to make an even more tailored media specific for certain cells. Due to extensive research on the type B genome, there are multiple alterations to be made to the media for

attempting to culture them. The alterations should have 2 intentions: i) selecting for the growth of the specific organism ii) limiting the growth of other organisms that could confound culturing of the target organism. Type B is genomically capable of synthesizing almost all 20 proteinogenic amino acids. The only amino acid with a gap in the genomic pathway is L-serine, as there is no phosphoserine phosphatase encoded for in the genome. As a result, this would be one of a few amino acids supplied for in the media as opposed to all 20. The genome also contains 2 *putP* L-proline transporters that are believed to be specific for the utilization of L-proline as a carbon and nitrogen source. This may be important for metabolism, therefore it too could be necessary in the media. Although there is no genomic evidence that type B does this, the presence of these transporters suggests these epulos might use L-proline through an unknown mechanism for carbon and nitrogen. L-threonine would also be included because it appears that type B can utilize it as an energy source, and L-cysteine would still be added as it serves as a reducing agent. Type B has the potential to make 3 B-vitamins, riboflavin, pyridoxal 5'-phosphate, and cobalamin. These vitamins would be excluded from the media as part of the vitamin supplement. *p*-amino benzoic acid would also be excluded, as they lack the genomic potential to utilize this compound to make folate. Another potential vitamin alteration would be to add the precursors of thiamin to the media as well, as it appears as if type B can salvage thiamin moieties. Type B also appears to utilize NO_3^- as an electron sink. The media contained some NO_3^- in the form of ammonia nitrate, but it is possible that not enough was present. Using more NO_3^- salts might be beneficial to growing and maintaining these cells. Type B has the genomic potential to utilize urea as a nitrogen source, adding only urea as a nitrogen source, with no added NH_3 , could be another selective measure.

Type B is quite versatile in its ability to utilize multiple complex and simple carbohydrates. The media used contained multiple sugars they have the genomic potential to assimilate, but other bacteria in the gut likely have the ability to assimilate these sugars as well. Limiting the diversity of sugars may limit the number of other community members from growing. Type B appears to be especially geared for galactose metabolism due to the high prevalence of galactans present in the host diet. Media could be made with D-galactose or L-galactose being the only carbon source to see if it could select for type B growth. Further, potentially harder to break down versions of galactose could be added as the sole carbohydrate. Type B has the genomic capacity to assimilate 3,6-anhydro-L-galactose from agarose and the D isomer from carrageenans. Other community members may lack this ability making this carbohydrate more selective for enriching for type B growth. This can be tried with only using one simple sugar as the carbon source. D-mannitol might be a good selective carbon source for type C and J cells as well as type Bs. Uronic acids could also potentially limit the growth of other organisms selected for epulo growth. Other methods could be employed to limit growth of other community members that may be confounding growth of epulos. Cells could be filtered through a large membrane (20 μ M), so only large morphotypes get stuck to the membrane. The membrane could then be washed and added directly to sterile media to select for the growth of the larger morphotypes. To potentially limit the growth of other spore formers in growth experiments from spore stocks, shortly after the stocks are added to media, they could be heated at high heats to kill all vegetative cells, which would likely not have had enough time to undergo a round of sporulation. Due to the lack of consistent germination of epulo spores, the epulo spores would be fine, while the other bacteria that germinate faster would die. The culture could then be subcultured and serially diluted to enrich for just epulo spores.

Another possibility is only providing complex carbohydrates to the media. Type B has the ability to degrade agars, porphyrans, mannans, carrageenans, xyloglucans, laminarin, and other complex algal polysaccharides. By providing these complex polysaccharides it may give type B a growth advantage in comparison to other community members that only utilize the soluble sugars generated from the degradation of these polysaccharides by other community members. This may require the culture to be done in serial dilutions because degradation of complex polysaccharides by epulos would possibly support the growth of other community members. If the cells were serially diluted to a lower inoculum, this may mitigate the problem of other community members overgrowing the culture. It is unclear what the best method of delivery of this should be. Purified compounds can be used to make semi-solid media, or solid media can be made in Hungate tubes, and the hydrolysis of these compounds could be observed. Although the algal extract did not appear to boost any potential growth in the media, sheets of dried algae could be sterilized and supplied as the only carbon source. The extraction process may create too many free sugars that would support the growth of other community members, but intact chunks of algae would likely maintain the complex polysaccharide nature of the compounds. Solid media could also be tried to culture type B, however, agar should not be used as a gelling agent. Rather a compound like gellan gum could be used instead, as the presence of agarases would break down the agar in the plates. Samples of the food bolus could also be collected from the gut contents, added to H₂O or buffer, autoclaved, made anaerobic, and then used as a media for growth. It appears as if type B might burrow into the food bolus and may like to have chunks of algae to attach to during polysaccharide degradation. This may also provide the cultures with potential host-derived factors that may be important for growth. Collecting known algal species the host fish feed on for each morphotype, could be used as a polysaccharide source as well.

These could either be ground up or added directly as chunks in the media. Inoculated tubes could be viewed for their degradation of the algae present in the media. Extract could also be made through chemical and heat digestion of the algae.

Type B also appears to be able to ferment citrate to generate oxaloacetate, which it can then utilize to generate a sodium motive force for energy. The original media contained citrate, but no other media did after that. This compound could be added as the sole carbon source, or along with one carbohydrate to enhance oxaloacetate production. Oxaloacetate could also be added to the media. Addition of these compounds should in turn increase the ability of type B to generate a sodium motive force. Na^+ content of the media may need to be adjusted to account for the SMF used for energy conservation. It could be that the media now has too high an external Na^+ gradient, preventing SMF generation. It is unclear if the type B cells utilize Na^+ or H^+ for powering their flagella, but adjustments to pH and Na^+ concentrations may be needed for motility and ATP synthesis through SMF.

B.11 REFERENCES

1. Clements K, Sutton D, Choat J. 1989. Occurrence and characteristics of unusual protistan symbionts from surgeonfishes (Acanthuridae) of the Great Barrier Reef, Australia. *Marine Biology* 102:403-412.
2. Angert ER, Clements KD, Pace NR. 1993. The largest bacterium. *Nature* 362:239-241.
3. Angert ER. 2012. DNA replication and genomic architecture of very large bacteria. *Annual review of microbiology* 66:197-212.
4. Clements K, Bullivant S. 1991. An unusual symbiont from the gut of surgeonfishes may be the largest known prokaryote. *Journal of bacteriology* 173:5359-5362.
5. Flint JF, Drzymalski D, Montgomery WL, Southam G, Angert ER. 2005. Nocturnal production of endospores in natural populations of Epulopiscium-like surgeonfish symbionts. *Journal of bacteriology* 187:7460-7470.
6. Angert ER, Clements KD. 2004. Initiation of intracellular offspring in Epulopiscium. *Molecular microbiology* 51:827-835.
7. Miller DA, Suen G, Clements KD, Angert ER. 2012. The genomic basis for the evolution of a novel form of cellular reproduction in the bacterium Epulopiscium. *BMC genomics* 13:265.
8. Miller DA, Choat JH, Clements KD, Angert ER. 2011. The spoIIE homolog of Epulopiscium sp. type B is expressed early in intracellular offspring development. *Journal of bacteriology* 193:2642-2646.

9. Tan IS, Ramamurthi KS. 2014. Spore formation in *Bacillus subtilis*. *Environmental microbiology reports* 6:212-225.
10. Haraga A, West TE, Brittnacher MJ, Skerrett SJ, Miller SI. 2008. *Burkholderia thailandensis* as a model system for the study of the virulence-associated type III secretion system of *Burkholderia pseudomallei*. *Infection and immunity* 76:5402-5411.
11. Fishelson L, Montgomery WL, Myrberg Jr AA. 1985. A unique symbiosis in the gut of tropical herbivorous surgeonfish(Acanthuridae: Teleostei) from the Red Sea. *Science(Washington)* 229:49-51.
12. Montgomery W, POLLAK PE. 1988. *Epulopiscium fishelsoni* ng, n. sp., a protist of uncertain taxonomic affinities from the gut of an herbivorous reef fish. *Journal of Eukaryotic Microbiology* 35:565-569.
13. Setlow P. 2003. Spore germination. *Current opinion in microbiology* 6:550-556.